

1 R. Joseph Trojan CA Bar No. 137,067  
trojan@trojanlawoffices.com  
2 TROJAN LAW OFFICES  
9250 Wilshire Blvd., Suite 325  
3 Beverly Hills, CA 90212  
Telephone: (310) 777-8399  
4 Facsimile: (310) 777-8348

5 Attorneys for Plaintiff,  
6 PHOENIX SOLUTIONS, INC.

7 **UNITED STATES DISTRICT COURT**

8 **NORTHERN DISTRICT OF CALIFORNIA**

9 **SAN FRANCISCO**

10  
11 PHOENIX SOLUTIONS, INC., a  
12 California corporation,

13 Plaintiff,

14  
15 v.

16  
17 WELLS FARGO BANK, N.A., a  
18 Delaware corporation,

19 Defendant.  
20

CASE NO. CV08-0863 MHP

**PLAINTIFF PHOENIX SOLUTIONS,  
INC.'S NOTICE OF MOTION AND  
MOTION TO STRIKE WELLS  
FARGO'S AFFIRMATIVE DEFENSE  
NUMBERS 35, 36, 37 AND 38;  
MEMORANDUM OF POINTS AND  
AUTHORITIES IN SUPPORT  
THEREOF**

21 DATE: July 14, 2008

22 TIME: 2:00 p.m.

23 JUDGE: MARILYN H. PATEL  
24  
25

26 //

27 //

28 **PLAINTIFF'S MOTION TO STRIKE**

**CV08-0863 MHP**

## TABLE OF CONTENTS

I.	INTRODUCTION .....	2
II.	STATEMENT OF FACTS .....	2
III.	ARGUMENT	
A.	Defendant's Affirmative Defenses did not meet the heightened pleading requirement required to plead inequitable conduct because Defendant did not plead sufficient facts to support the elements of inequitable conduct .....	4
1)	Defendant failed to adequately plead relevant and undisclosed prior arts because Defendant did not sufficiently state how the Prior-Art-Patents were material to the Patents-in-Suit's prosecution.....	4
a.	Allegations against the '846 Patent based on alleged omissions for the Stanford and Trower Patents .....	6
b.	Allegations against the '640 Patent based on alleged omissions for Flanagan, Haddock and Chou Patents .....	7
c.	Allegations against the '977 Patent based on alleged omissions for Waters, Flanagan, Haddock, Chou, Stanford and Trower Patents .....	9
d.	Allegations against the '854 Patent based on alleged omissions for Giangarra, Horiguchi, Kuhn and Trower Patents.....	10
2)	Even assuming Defendant sufficiently stated how the Prior-Art-Patents were relevant prior art, Defendant failed to adequately plead how omission of the Prior-Art-Patents were meant to mislead the USPTO .....	12

1 B. If the Court is inclined to strike Defendant's inequitable  
2 conduct affirmative defenses, Phoenix requests the Court not  
3 allow Defendant leave to amend because Defendant cannot, in  
4 good faith, show how the Prior-Art-Patents were material to the  
5 Patents-in-Suit's prosecution..... 14

6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28 IV. CONCLUSION..... 14

TROJAN LAW OFFICES  
BEVERLY HILLS

TABLE OF AUTHORITIES

CASES

<u>Cargill, Inc. v. Canbra Foods, Ltd.</u> , 476 F.3d 1359 (Fed. Cir. 2007).....	4
<u>Central Admixture Pharmacy Services, Inc. v. Advanced Cardiac Solutions</u> , 482 F.3d 1347 (Fed. Cir. 2007).....	4, 5, 12, 13
<u>Digital Control, Inc. v. Charles Mach. Works</u> , 437 F.3d 1309 (Fed. Cir. 2006).....	5
<u>Ferguson Beauread/Logic Controls, Inc. v. Mega Sys., LLC</u> , 350 F.3d 1327 (Fed. Cir. 2003).....	2, 4

CODES, RULES, AND STATUTES

35 U.S.C. § 271 .....	2
Fed.R.Civ.P. 12(f) .....	1, 14
37 Code of Federal Regulation § 1.56 .....	5, 6

NOTICE OF MOTION AND MOTION

PLEASE TAKE NOTICE, that on July 14, 2008, at 2:00 p.m. before the Honorable Marilyn H. Patel, Plaintiff PHOENIX SOLUTIONS, INC. ("Phoenix") will, and hereby does, move the Court for an order striking Defendant, WELLS FARGO BANK, N.A.'s ("Wells Fargo") affirmative defense numbers 35, 36, 37 and 38 pursuant to Federal Rules of Civil Procedure, Rule 12(f).

This motion is based upon this Notice of Motion and Motion; a Memorandum of Points and Authorities below; the Declaration of R. Joseph Trojan; all papers and pleadings filed herein; and on such other matters as may be allowed by the Court.

ISSUES TO BE DECIDED (Civil Local Rule 7-4(a)(3))

1. Whether Wells Fargo's affirmative defenses of inequitable conduct should be stricken?

## MEMORANDUM OF POINTS AND AUTHORITIES

### I. INTRODUCTION

An allegation of inequitable conduct calls for a heightened pleading standard and must be plead with particularity. Ferguson Beauread/Logic Controls, Inc. v. Mega Sys., LLC, 350 F.3d 1327, 1344 (Fed. Cir. 2003). Here, Defendant has alleged inequitable conduct without pleading sufficient facts for Plaintiff to prepare an adequate response. Defendant has pleaded that Plaintiff should have disclosed references cited to Plaintiff in patent applications claiming different subject matter. Notably, Defendant makes a feeble and confusing attempt to show that these references were material by selecting a few choice words from the reference and finding similar words in the claims of the Patents-at-Suit. However, Defendant did not state how these references were material or even relevant. Further, Defendant did not state how omissions of these references were meant to mislead the USPTO. Without more, Plaintiff is left to guess as to the basis of Defendant's inequitable conduct defense. Thus, Defendant's defenses were not pleaded with sufficient particularity. Therefore, the Court should strike the Defendant's affirmative defense numbers 35, 36, 37 and 38.

### II. STATEMENT OF FACTS

Phoenix Solutions, Inc. ("Plaintiff" or "Phoenix"), sued Wells Fargo, N.A. ("Defendant" or "Wells Fargo"), for patent infringement under 35 U.S.C. § 271 *et seq.* See Exh. 1, Declaration of R. Joseph Trojan ("Trojan Decl."). The patents at issue are U.S. Patent Nos. 6,633,846 ("846 Patent"), 6,665,640 ("640 Patent"), 7,050,977 ("977 Patent") and 7,277,854 ("854 Patent") (Collectively "Patents-in-Suit"). The Patents-in-Suit are directed to speech recognition software. Speech recognition software allows users to verbally communicate with a computer by

1 allowing the computer to process the user's speech and then generate appropriate  
2 responses. *See* Exh.1, Trojan Decl.

3 Defendant's affirmative defense numbers 35, 36, 37 and 38 each contend that  
4 the Patents-in-Suit are unenforceable due to Plaintiff's inequitable conduct before  
5 the United States Patent and Trademark Office ("USPTO"). *See* Exh. 2, Trojan  
6 Decl. (¶¶82-147).

7 Specifically, Defendant alleges that Plaintiff violated the duty of candor for  
8 not disclosing prior art patents during the application process of the Patents-in-Suit.  
9 *See* Exh.2, Trojan Decl. (¶¶82-147). For the '846 Patent, Defendant alleges the  
10 following references were improperly omitted: 5,615,296 to Stanford ("Stanford  
11 Patent") and 5,983,190 to Trower ("Trower Patent"). For the '640 Patent,  
12 Defendant alleges the following references were improperly omitted: 5,737,485 to  
13 Flanagan ("Flanagan patent"); 5,265,014 to Haddock ("Haddock Patent"); and  
14 6,336,090 to Chou ("Chou Patent"). For the '977 Patent, Defendant alleges the  
15 following references were improperly omitted: Stanford Patent; Flanagan patent;  
16 Haddock Patent; 5,540,589 to Waters ("Waters Patent"); Chou Patent; and Trower  
17 Patent. For the '854 Patent, Defendant alleges the following references were  
18 improperly omitted: Trower Patent; 6,101,472 to Giangarra ("Giangarra Patent");  
19 6,330,530 to Horiguchi ("Horiguchi Patent"); and 6,901,366 to Kuhn ("Kuhn  
20 Patent") (Collectively "Prior-Art-Patents").

### 21 22 III. ARGUMENT

23 Defendant's "facts" in support of their defense consist of nothing more than  
24 comparing the lists of which references were cited in each application to find any  
25 minor differences, and then cherry picking a few words from a reference out of  
26 context. Defendant attempts to magnify such language into a claim for inequitable  
27 conduct. Other than pointing out a few words out of context, Defendant's pleading

1 lacks any explanation or contention that any of such references were in fact material  
2 to the claims of the Patents-in-Suit. Thus, as a matter of law, it cannot meet the  
3 heightened pleading requirements for this defense.

4  
5 A. Defendant's Affirmative Defenses did not meet the heightened  
6 pleading requirement required to plead inequitable conduct because  
7 Defendant did not plead sufficient facts to support the elements of  
8 inequitable conduct.

9 To prove inequitable conduct, defendant must show by clear and convincing  
10 evidence that the patent applicant (1) either "made an affirmative misrepresentation  
11 of material fact, failed to disclose material information, or submitted false material  
12 information, and (2) intended to deceive the [USPTO]." Cargill, Inc. v. Canbra  
13 Foods, Ltd., 476 F.3d 1359, 1363 (Fed. Cir. 2007). The Federal Circuit has held  
14 that "inequitable conduct, while a broader concept than fraud, must be pled with  
15 particularity." Central Admixture Pharmacy Services, Inc. v. Advanced Cardiac  
16 Solutions, 482 F.3d 1347, 1356 (Fed. Cir. 2007)(quoting Ferguson  
17 Beauregard/Logic Controls v. Mega Systems, 350 F.3d 1327, 1344 (Fed. Cir.  
18 2003)).

19 To meet the higher pleading standard, a defendant must plead facts to show  
20 (1) what relevant and undisclosed prior art was known to the patentee and (2) how  
21 patentee's actions were meant to mislead the USPTO. Central Admixture  
22 Pharmacy Services, Inc., 482 F.3d at 1356-1357.

23  
24 1) Defendant failed to adequately plead relevant and undisclosed  
25 prior art because Defendant did not sufficiently state how the  
26 Prior-Art-Patents were material to the Patents-in-Suit's  
27 prosecution.

28 In pleading inequitable conduct, defendant must first plead facts to show



1 what relevant and undisclosed prior art was known to the patentee in order “to give  
2 notice to the other party of the facts on which the defense is premised.” Central  
3 Admixture Pharmacy Services, Inc., at 1356-1357. Under the “reasonable  
4 examiner” standard, relevant prior art is prior art a reasonable examiner would have  
5 considered important in deciding whether to allow the application. Digital Control,  
6 Inc. v. Charles Mach. Works, 437 F.3d 1309, 1314 (Fed. Cir. 2006).

7 Patent applicants have a duty to disclose information material to  
8 patentability. See 37 Code of Federal Regulation (“CFR”) section 1.56, emphasis  
9 added. Information is material:

10 When it is not cumulative to information already of record or  
11 being made of record in the application, and (1) It establishes,  
12 by itself or in combination with other information, a prima facie  
13 case of unpatentability of a claim; or (2) It refutes, or is  
14 inconsistent with, a position the applicant takes in: (i) Opposing  
15 an argument of unpatentability relied on by the Office, or (ii)  
16 Asserting an argument of patentability.

17 37 CFR Section 1.56(b). Further, “a withheld otherwise material prior art reference  
18 is not material for the purposes of inequitable conduct if it is merely cumulative to  
19 that information considered by the examiner.” Digital Control, Inc. v. Charles  
20 Mach. Works, 437 F.3d 1309, 1319 (Fed. Cir. 2006).

21 The fact that Defendant can find some limitation in the claim, and then argue  
22 that a reference has some teaching relative to that is not the test for materiality. See  
23 37 C.F.R. section 1.56(b), supra. By Defendant’s definition, there is no such thing  
24 as a non-material reference. Materiality must be related in context to the point of  
25 distinction in the patented claims. Defendant merely claims the Prior-Art-Patents  
26 must be material because it was cited against a patent in the “same family” without  
27 reference to what claim was involved, or why the reference was cited. Defendant  
28 has steered clear from any direct statement of reasons why it would be material to

1 any particular claim.

2 Further, as required by the definition of material information under 37 CFR  
3 section 1.56(b), Defendant has not explained how these references are not  
4 cumulative. As discussed below, Defendant pleaded references that were  
5 cumulative of the references that were already cited to the patent examiner. In fact,  
6 an Examiner would not care about the omitted references because he already had a  
7 better reference in hand for the claim Defendant is attacking. Thus, such references  
8 were not material information to the patent application and Phoenix did not have a  
9 duty to disclose them. Without factual pleadings to why such references were not  
10 cumulative, Phoenix cannot prepare an adequate response.

11  
12 a. Allegations against the '846 Patent based on alleged  
13 omissions for the Stanford and Trower Patents.

14 Defendant alleges the Stanford Patent was material to the '846 Patent  
15 because the '846 Patent claims a system "wherein said speech representatives  
16 values are transmitted continuously during said speech utterances" and the Stanford  
17 Patent discloses a "technique of speaker-independent, continuous-speech phrases  
18 and bi-grams." See Exh. 2, Trojan Decl. (¶ 84).

19 Defendant has taken the excerpt of the '846 Patent from claim 1 of the  
20 patent. However, during the prosecution of the case, the Examiner had already  
21 cited a much more relevant reference (Barclay et al. – which the Applicant had  
22 disclosed to the Examiner) against then pending claim 1. In fact, the Examiner  
23 claimed that Barclay taught all the limitations of then pending claim 1, including  
24 the limitation cited by Defendant here.

25 In short, Defendant has not provided any facts to explain why the Stanford  
26 Patent is not cumulative to the other prior art of record. Without a factual pleading  
27 that explains why the Stanford Patent would have been considered more relevant to

1 the claims than the prior art of record, Phoenix is only left to guess as to  
2 Defendant's basis for its defense.

3 Further, Defendant alleges the Trower Patent was material to the '846 Patent  
4 because the '846 Patent claims a program used in a system "for receiving user  
5 speech utterance signals representing speech utterances to be recognized" that  
6 "works within a browser program executing on said computing system" and the  
7 Trower Patent discloses a system that relates to "speech input" and utilizes "a  
8 microphone and analog to digital convertor circuitry for converting sound to  
9 digitized audio" and that the system is "advantageous for web pages." See Exh. 2,  
10 Trojan Decl. (¶ 89).

11 Here, Defendant's pleading is defective for the same reasons as the Stanford  
12 Patent. Defendant has not pleaded how the Trower reference is not cumulative to  
13 the prior art of record. Further, the "works within a browser program" language  
14 Defendant has quoted comes from dependent claim 2 of the '846 Patent and it is the  
15 only claim in the '846 Patent which makes a reference to the "browser" language  
16 Defendant cites. For a reference to be non-cumulative and material it must be  
17 relevant to the patentability of claim 1. Thus, Defendant has only alleged that the  
18 Trower reference shows "receiving user speech utterance signals." There are  
19 multiple references cited that also show this same limitation. Therefore, Plaintiff  
20 does not know how such reference is material to the claims of the '846 Patent.

21  
22 b. Allegations against the '640 Patent based on alleged  
23 omissions for Flanagan, Haddock and Chou Patents.

24 Defendant alleges the Flanagan Patent was material to the '640 Patent  
25 because the '640 Patent claims "a speech recognition system for generating  
26 recognized speech utterance data from partially processed speech data" and the  
27 Flanagan Patent discloses a "feature extractor [that] extracts speech features or

1 cepstrum coefficients,” which data are then “provided as inputs to the speech  
2 recognizer.” *See* Exh. 2, Trojan Decl. (¶ 95).

3 The Flanagan Patent was cited entirely for the cepstrum coefficients.  
4 Phoenix cited nearly a dozen different separate references for that concept. Again,  
5 the Flanagan Patent adds nothing to any of these. Moreover, during the ‘640 Patent  
6 prosecution, the Examiner cited the Stanford Patent discussed above, as anticipating  
7 claim 1 as filed. Thus, the question becomes, how is the Flanagan reference better  
8 than the Stanford reference applied by the Examiner? Defendant’s facts only allege  
9 that the Flanagan Patent is germane to one limitation in the claim. The Examiner  
10 was considering references that he believed disclosed all the limitations in the  
11 claims.

12 Further, Defendant alleges the Haddock Patent was material to the ‘640  
13 Patent because the ‘640 Patent claims a system “adapted for responding to speech-  
14 based queries” that has a “speech recognition system for generating recognized  
15 speech utterance data” and “a query formulation system for converting said  
16 recognized speech data into a search query suitable for identifying a topic query  
17 entry corresponding to said speech-based query” and the Haddock Patent discloses  
18 a system whereby “the user communicates textual information to the computer  
19 system by talking to the computer rather than by typing the information at the  
20 keyboard” and is “embodied in a user interface of a database system which receives  
21 a database query from a user, evaluates the query, and provides a result of the  
22 evaluation to the user.” *See* Exh. 2, Trojan Decl. (¶ 99).

23 Here, the Haddock Patent is generally argued to be a reference that shows  
24 talking to a computer and using a database. Again, Defendant failed to show the  
25 Haddock Patent is more relevant than the Stanford Patent, cited by the Examiner,  
26 which was argued to disclose all limitations of then pending claim 1. Defendant  
27 has not provided any facts that the Haddock Patent is non-cumulative.

Furthermore, Defendant alleges the Chou Patent was material to the '640 Patent because the '640 Patent claims a system that involves "partially processed speech data being received from a remote speech capturing system" and the Chou Patent discloses a "feature extraction and/or ASR units can be located at the receiving base station, the switch connect to the base station... or at another location connection on the network(s) to which these elements are connected" and that it will sometimes "be convenient to have the feature extraction and ASR operations performed at different locations." *See* Exh. 2, Trojan Decl. (¶ 103).

Once again, Defendant failed to show the Chou Patent is more relevant than the Stanford Patent, cited by the Examiner, which was argued to disclose all limitations of then pending claim 1. Phoenix submitted dozens of references on this type of limitation and thus disclosure of the Chou Patent would have been cumulative.

c. Allegations against the '977 Patent based on alleged omissions for Waters, Flanagan, Haddock, Chou, Stanford and Trower Patents.<sup>1</sup>

Defendant alleges the Waters Patent was material to the '977 Patent because the '977 Patent claims a system "wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively" and the Waters Patent discloses a system where the "voice recognizer is illustrated as a standalone component, although it may be built-in to the controller." *See* Exh. 2, Trojan Decl. (¶ 120).

The bulk of the claims in the '977 Patent were rejected in light of a reference

<sup>1</sup> The Flanagan, Haddock, Chou, Stanford and Trower Patents were cited here again for the same propositions as they were cited in the previous affirmative defenses. Please see discussion above.

1 to *White et al.* taken with other references. Nowhere does Defendant explain how  
2 the Waters Patent would be non-cumulative over White, the main reference cited in  
3 the file history. The Waters Patent was applied as a secondary reference to 3 minor  
4 dependent claims out of 70 in the '846 Patent Application on September 21, 2001.

5  
6 d. Allegations against the '854 Patent based on alleged  
7 omissions for Giangarra, Horiguchi, Kuhn and Trower  
8 Patents.<sup>2</sup>

9 Defendant alleges the Giangarra Patent was material to the '854 Patent  
10 because the '854 Patent claims a method that includes "providing a speech  
11 recognition engine adapted to recognize a first set of words and/or phrases during  
12 an interactive speech session" and the Giangarra Patent discloses a "vocabulary list  
13 stored in speech recognition unit [that] provides a list of all words and utterances by  
14 an external user which will be recognized as voice commands." See Exh. 2, Trojan  
15 Decl. (¶ 137).

16 Basically, the Giangarra Patent is being alleged as material because it  
17 discloses a grammar, a concept that is found in almost 100 other references. Of the  
18 26 claims in the '854 Patent, the only substantive rejection was of claim 19 based  
19 on an obviousness rejection using Harless (U.S. Patent No. 5,730,603) and Claassen  
20 (U.S. Patent No. 6,637,363). Nowhere does Defendant explain how the Giangarra  
21 Patent would be non-cumulative over Harless, the main reference cited in the file  
22 history. Again, Giangarra was applied as a secondary reference to 4 minor  
23 dependent claims out of 26 in the '977 on August 10, 2004 as disclosing an  
24 interactive lesson tutorial.

25 Further, Defendant alleges that the Horiguchi Patent was material to the '854  
26 Patent because the '854 Patent claims "a natural language query sytem" and the

27 <sup>2</sup> The Trower Patent was cited here again for the same proposition as it was cited in the previous  
28 affirmative defenses. Please see discussion above.

1 Horiguchi Patent describes a “natural language processing system.” *See* Exh. 2,  
2 Trojan Decl. (¶ 141).

3 For the ‘854 Patent, Phoenix cited an abundance of natural language  
4 processing systems references. The affirmative defense fails to explain how the  
5 Horiguchi Patent contributes over and above the other references that were used  
6 against the ‘854 claims. The Horiguchi Patent was applied as a secondary reference  
7 to 4 minor dependent claims out of 26 in the ‘977 Patent Application on August 19,  
8 2004, as a natural language system disclosing predetermined sentences used to  
9 determine a recognized sentence. It is only cited by Defendant against this single  
10 limitation of the ‘854 Patent, unlike the more relevant references cited in the ‘854  
11 file history.

12 Finally, Defendant alleges the Kuhn Patent was material to the ‘854 Patent  
13 because the ‘854 Patent claims a method of using a system that provides “a  
14 database of query/answer pairs concerning one or more topics which can be  
15 responded to by the natural language query system” and the Kuhn Patent discloses a  
16 “knowledge database” as well as a “natural language parser [that] analyzes and  
17 extracts semantically important and meaningful topics from a loosely structured,  
18 natural language text.” *See* Exh. 2, Trojan Decl. (¶ 145).

19 The Kuhn Patent cannot possibly be more relevant than Harless, cited as the  
20 main reference in the ‘854 Patent, and all the other cited references. Defendant do  
21 not allege it as so, either. The Kuhn Patent was applied as a secondary reference to  
22 a single (1) minor dependent claims out of 26 in the ‘977 Patent Application in June  
23 2005 as accessing info over the Internet by speech recognition using multiple  
24 speech engines. It is only cited by Defendant against this single “database...”  
25 limitation of the ‘854 Patent, unlike the more relevant references cited in the ‘854  
26 file history.

27 For all the references that Defendant claimed were wrongly omitted,  
28 **PLAINTIFF’S MOTION TO STRIKE** -11- **CV08-0863 MHP**



1 Defendant has not provided any facts to explain why these references were not  
2 cumulative to the other prior art of record. Without a factual pleading that explains  
3 why these references were not cumulative and why an Examiner would have  
4 considered them more relevant to the claims than the prior art of record, Phoenix is  
5 only left to guess as to Defendant's basis for its defense. Thus, Defendant has not  
6 sufficiently plead its affirmative defenses to meet the heightened pleading  
7 requirements for its inequitable conduct allegations.

8  
9 2) Even assuming Defendant sufficiently stated how the Prior-Art-  
10 Patents were relevant prior art, Defendant failed to adequately  
11 plead how omission of the Prior-Art-Patents were meant to  
12 mislead the USPTO.

13 In pleading inequitable conduct, defendant must also plead facts to show how  
14 a patentee's actions were meant to mislead the USPTO. Central Admixture  
15 Pharmacy Services, 482 F.3d at 1356-1357.

16 In Central Admixture Pharmacy Services, the defendants pled, "by  
17 manipulation of various measurements and units, the patentee sought to mislead the  
18 Patent and Trademark Office regarding the relationship between the claimed  
19 invention and the prior art." Id. at 1356 (internal citations omitted). The district  
20 court ruled that the pleading was insufficient to allege inequitable conduct. The  
21 Federal Circuit affirmed the district court's ruling, finding that defendants did not  
22 plead "**how** the manipulation was meant to mislead the PTO." Id. at 1357  
23 (emphasis added). The Federal Circuit distinguished between pleading facts  
24 showing **how** patentee's actions were meant to mislead and pleading facts **asserting**  
25 **patentee sought** to mislead. The court found that pleading facts asserting patentee  
26 sought to mislead, without pleading facts as to how, was insufficient. Id.

27 Here, Defendant pleads that Phoenix "withheld the information with the



1 intent to deceive the USPTO.” Then, Defendant went to great lengths to emphasize  
2 that Phoenix had multiple Information Disclosure Statements and claim  
3 amendments during the Patents-in-Suit prosecutions. However, Defendant did not  
4 state *how* not specifically disclosing the Prior Art Patents was meant to mislead the  
5 USPTO. For example, as to the ‘977 Patent, Defendant claimed Phoenix should  
6 have disclosed four references cited to Phoenix during the prosecution of the ‘846  
7 Patent. See Exh. 3, Trojan Decl. Defendant does not explain how the Examiner in  
8 the ‘977 Patent could be “misled” by a failure to disclose the four ‘846 references  
9 that he already knew about because he was simultaneously handling the ‘846  
10 application as well. If, as Defendant suggests the subject matter was so “related”  
11 one would expect a reasonable Examiner, knowing of such references, to apply the  
12 same from the ‘846 into the ‘977 as well. The lack of citation of these references  
13 by the same Examiner in the ‘977 in fact illustrates precisely how weak and  
14 inadequate Defendant’s allegations of materiality are in this case.

15 Finally, the fact that Phoenix had multiple Information Disclosure Statements  
16 and claim amendments during the prosecution of the Patents-in-Suit does not show  
17 how Phoenix’s actions were meant to deceive the USPTO. To the contrary,  
18 multiple Information Disclosure Statements and claim amendments are normal  
19 procedures in patent prosecution, and in fact illustrate that Phoenix’s patent  
20 attorney was acutely conscious and conscientious in bringing any relevant prior art  
21 – including some 100+ references over the course of six and a half years to the  
22 Examiner’s attention for the ‘977 Patent application.

23 The present pleading of inequitable conduct is clearly defective as is, and, at  
24 a minimum Phoenix cannot decipher what facts Defendant’s affirmative defense is  
25 premised on. Thus, as in Central Admixture Pharmacy Services, this court should  
26 strike Defendant’s pleading for insufficiently pleading inequitable conduct with  
27 particularity.

1 B. If the Court is inclined to strike Defendant's inequitable conduct  
2 affirmative defenses, Phoenix requests the Court not allow Defendant  
3 leave to amend because Defendant cannot, in good faith, show how the  
4 Prior-Art-Patents were material to the Patents-in-Suit's prosecution.

5 As discussed above, Defendant cannot show how the Prior-Art-Patents were  
6 material prior art to the Patents-in-Suit. Thus, if the court is inclined to strike  
7 Defendant's affirmative defense numbers 35, 36, 37 and 38, Phoenix requests the  
8 court not allow Defendant leave to amend.

9 IV. CONCLUSION

10 For the above reasons, Phoenix request Defendant's affirmative defense  
11 numbers 35, 36, 37 and 38 be stricken pursuant to Federal Rules of Civil Procedure,  
12 Rule 12(f). Further, Phoenix requests, if the court is inclined to strike Defendant's  
13 affirmative defense numbers 35, 36, 37 and 38, the court not allow Defendant leave  
14 to amend.

15 //

16 //

17 //

18 //

19 //

20  
21  
22 TROJAN LAW OFFICES  
23 By:

24 Date: May 27, 2008

25 /s/R. Joseph Trojan  
26 R. Joseph Trojan  
27 Attorney for Plaintiff  
Phoenix Solutions, Inc.

R. Joseph Trojan CA Bar No. 137,067  
trojan@trojanlawoffices.com  
TROJAN LAW OFFICES  
9250 Wilshire Blvd., Suite 325  
Beverly Hills, CA 90212  
Telephone: (310) 777-8399  
Facsimile: (310) 777-8348

Attorneys for Plaintiff,  
PHOENIX SOLUTIONS, INC.

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO**

PHOENIX SOLUTIONS, INC., a  
California corporation,

Plaintiff,

v.

WELLS FARGO BANK, N.A., a  
Delaware corporation,

Defendant.

CASE NO. CV08-0863 MHP

**DECLARATION OF R. JOSEPH  
TROJAN IN SUPPORT OF  
PLAINTIFF PHOENIX SOLUTIONS,  
INC.'S MOTION TO STRIKE  
DEFENDANT'S AFFIRMATIVE  
DEFENSE NUMBERS 35, 36, 37 AND  
38**

DATE: July 14, 2008

TIME: 2:00 p.m.

JUDGE: MARILYN H. PATEL

//

//

//

**DECLARATION OF R. JOSEPH TROJAN**

**CV08-0863 MHP**

TROJAN LAW OFFICES  
BEVERLY HILLS

1 I, R. Joseph Trojan, declare as follows:

2 1. I am an attorney at law, duly licensed to practice law in the State of  
3 California. I am the Principal of the Trojan Law Offices, the attorneys of record for  
4 Plaintiff, Phoenix Solutions, Inc. I have personal knowledge of the facts stated  
5 herein. If called upon to do so, I could and would competently testify that:  
6

7 2. Attached hereto as **Exhibit 1** is a true and correct copy of Phoenix  
8 Solutions, Inc.'s ("Phoenix") Amended Complaint for Patent Infringement (original  
9 exhibits omitted).  
10

11 3. Attached hereto as **Exhibit 2** is a true and correct copy of Wells Fargo,  
12 N.A.'s Answer to Amended Complaint.  
13

14 4. Attached hereto as **Exhibit 3** is a true and correct copy of U.S. Patent  
15 No. 7,050,977 (original exhibit 3 to Phoenix's Amended Complaint).  
16

17 I declare under penalty of perjury under the laws of the United States that the  
18 foregoing is true and correct. Executed on May 27, 2008, in Beverly Hills,  
19 California.  
20

21  
22  
23 /s/R. Joseph Trojan  
24 R. Joseph Trojan  
25  
26  
27

# EXHIBIT 1

TROJAN LAW OFFICES  
BEVERLY HILLS

1 R. Joseph Trojan CA Bar No. 137,067  
2 trojan@trojanlawoffices.com  
3 TROJAN LAW OFFICES  
4 9250 Wilshire Blvd., Suite 325  
5 Beverly Hills, CA 90212  
6 Telephone: 310-777-8399  
7 Facsimile: 310-777-8348

8 Attorneys for Plaintiff,  
9 PHOENIX SOLUTIONS, INC.

10 **UNITED STATES DISTRICT COURT**  
11 **NORTHERN DISTRICT OF CALIFORNIA**  
12 **SAN FRANCISCO DIVISION**

13 PHOENIX SOLUTIONS, INC., a  
14 California corporation,

15 Plaintiff,

16 v.

17 WELLS FARGO BANK, N.A., a  
18 Delaware corporation,

19 Defendant.

CASE NO. CV08-00863MHP

**AMENDED COMPLAINT FOR  
DAMAGES AND INJUNCTIVE  
RELIEF FOR INFRINGEMENT OF  
U.S. PATENT NOS. 6,633,846,  
6,665,640, 7,050,977 AND 7,277,854  
UNDER 35 U.S.C. § 271 AND  
DEMAND FOR JURY TRIAL  
PURSUANT TO FED. R. CIV.  
PROC., RULE 38**

20 //

21 //

22 //

23 //

24 //

25 //

26 //

27 //

28 **AMENDED COMPLAINT**

**CASE NO. CV08-0863 MHP**

1 Plaintiff, PHOENIX SOLUTIONS, INC. (hereinafter "Plaintiff" or  
2 "Phoenix"), hereby complains against Defendant, WELLS FARGO BANK, N.A.  
3 (hereinafter "Defendant" or "Wells Fargo"), as follows:

4 1. This is a civil action for patent infringement arising under the patent  
5 laws of the United States, 35 U.S.C. § 271 *et seq.*

### 6 **I. THE PARTIES**

7 2. Plaintiff is a corporation organized and existing under the laws of the  
8 State of California, with a place of business at 634 Georgia Avenue, Palo Alto,  
9 California, 94306.

10 3. Upon information and belief, Defendant is a corporation organized and  
11 existing under the laws of the State of Delaware with a place of business at 420  
12 Montgomery Street, San Francisco, California, 94163.

### 13 **II. FACTUAL BACKGROUND**

14 4. Plaintiff is the owner by assignment of U.S. Patent Nos. 6,633,846,  
15 6,665,640, 7,050,977, and 7,277,854 (hereinafter "Patents in Suit") directed to  
16 "speech recognition software".

17 5. Plaintiff Phoenix developed the next generation of speech recognition  
18 systems that give users the ability to have a verbal conversation with a computer  
19 about a subject on which the computer has been programmed to process and  
20 generate intelligent responses. One of the first applications of this new technology  
21 was its use in telephone customer service lines where the customer calls a computer  
22 and a "virtual customer service agent" answers the line and interacts with a caller  
23 using "natural speech" akin to a live person.

24 6. Phoenix encompasses the life work of a pioneer in the field of  
25 computer-based speech recognition, Dr. Ian Bennett. Originally from Jamaica, Dr.  
26 Bennett graduated with honors from the University of British Columbia and went  
27 on to receive his Master's and Doctorate degrees in electrical engineering from

1 Stanford University. While at Stanford, Dr. Bennett developed the first practical  
2 analog processor for speech compression. After graduation he held technical  
3 engineering positions with several high technology companies and contributed to  
4 device and product development. As a consultant to the Variable Speech Corp. of  
5 Tokyo, Japan, he contributed to the development of an analog speech compression  
6 VLSI speech processor used for audio compression in consumer speech recorders.  
7 In 1994, Dr. Bennett began the development of a natural language query system  
8 (NLQS). Subsequently, he founded Phoenix Solutions, where he guided the  
9 development of algorithms for statistics- and semantics-based signal processing of  
10 speech that allow a computer to take in natural speech questions and return answers  
11 that also sound like natural speech. Dr. Bennett developed various applications for  
12 his technology, including interactive conversational systems and interactive guides,  
13 intelligent tutoring systems and form-filling systems. Dr. Bennett is currently at the  
14 National Science Foundation serving as a Program Director within the Directorate  
15 of Engineering, Division of Industrial Innovation & Partnerships.

16 7. Defendant Wells Fargo is a financial services company that provides  
17 banking, insurance, investment, mortgage loan, and consumer finance services. In  
18 connection with its electronic services, Defendant (and/or others on its behalf)  
19 established and operates a number of customer support lines, which can be reached  
20 for example at (800) 642-4720 and upon information and belief, other toll-free  
21 phone numbers. The customer support lines employ a natural language interactive  
22 voice response (IVR) system that includes a virtual agent (hereinafter  
23 interchangeably referred to as "IVR system").

24 8. The Plaintiff's natural language IVR system is superior to  
25 conventional touch-tone systems because the caller can simply talk to the system  
26 using natural language. In contrast, touch-tone IVR systems require the caller to  
27 select from a series of choices using a more limited telephone keypad. IVR touch



1 tone systems are also less efficient since they require callers to listen to an entire  
2 menu of choices and wade through a series of menus before providing a response to  
3 the caller. Consumers hang up at a greater rate in frustration when they become  
4 lost in the maze of menus.

5 9. The alternative to touch tone menu systems is to employ live  
6 operators. When compared to live operators, the Plaintiff's IVR system is much  
7 more cost effective. Based upon industry data, it is estimated that Defendant's use  
8 of its current IVR system has allowed it to save 93% of the cost it previously  
9 incurred in providing its customer support line and Defendant's customer  
10 satisfaction has increased by 30%.

11 10. Upon information and belief, Defendant operates its IVR system using  
12 a combination of telephony hardware and computer server hardware that is  
13 specifically adapted by Defendant (and/or others on its behalf) to respond to spoken  
14 questions from callers concerning the Defendant's business. Such hardware uses  
15 supporting software that includes speech recognition and natural language engines  
16 used to understand the spoken questions from callers.

17 11. Upon information and belief, the speech recognition engine used by  
18 Defendant is distributed, so that some of the speech-processing operations for  
19 understanding callers are performed on a client computing system (such as  
20 telephony platform or other hardware) while other speech processing operations are  
21 performed on a separate server computing system. Upon information and belief,  
22 Defendant (and/or others on its behalf) configure such computing systems to  
23 customize what speech processing operations will take place on such respective  
24 hardware systems to maximize certain characteristics of the system, and to regulate  
25 how speech data from the callers is transferred between such systems.

26 12. When customers place calls to Defendant's IVR system, they can  
27 speak in a conversational style as if they were speaking to a real person.

1 Defendant's interactive virtual agent responds to the caller's questions in real-time  
2 by providing answers in natural speech. The virtual agent has been taught natural  
3 language dialogues based on information concerning Defendant's products  
4 provided by the Defendant and incorporated into the software. In this manner, the  
5 virtual agent can understand questions posed by customers concerning Defendant's  
6 products, and give relevant answers.

7 13. Defendant's IVR system uses a speech recognition engine to break  
8 down the customer's questions into specific words understood by the IVR system.  
9 For example, the speech recognition engine could determine that the user has said  
10 his or her account number. Defendant controls precisely what specific words its  
11 IVR system will understand as part of its vocabulary by configuring (and/or having  
12 others configure on its behalf) certain aspects of such client computing system  
13 and/or server computing system.

14 14. Defendant's IVR system employs a natural language engine to  
15 understand the meaning of the specific words spoken by its customers. The IVR  
16 system, by understanding the meaning and context of specific words, may  
17 determine that the customer is asking about a service related problem. Defendant  
18 controls precisely what interpretation the IVR system should give to various words  
19 spoken by its customers by configuring (and/or having others configure on its  
20 behalf) certain aspects of the client computing system and/or server computing  
21 system.

22 15. Based on determining the most likely meaning of the customer's  
23 specific question, the interactive virtual agent responds with a specific answer. The  
24 answer may take the form of an audible response from the agent, or it may take the  
25 form of the IVR system routing the caller to a live person working within the  
26 appropriate department (such as the service department in the example above). In  
27 all instances, Defendant alone controls precisely what responses and actions virtual

1 agent takes, and has configured (and/or has had others configure on its behalf)  
2 certain aspects of such client computing system and/or server computing system to  
3 provide such desired responses or actions.

4 16. Upon information and belief, Defendant also configured and controlled  
5 (and/or has had others configure and/or control on its behalf) other aspects of the  
6 virtual agent's overall behavior, including among other things, the gender, apparent  
7 age, speech rate, prosody, style and rate of response. These parameters are selected  
8 and controlled by Defendant to increase customer satisfaction with the customer  
9 support line.

10 17. Upon information and belief, Defendant (and/or others on its behalf)  
11 designed, customized and selected the personality exhibited by the virtual agent as  
12 well. This electronic persona was specifically selected to be appealing and  
13 attractive to Defendant's customers and to maximize utilization of the IVR system  
14 by such customers.

15 18. Upon information and belief, the information used by Defendant's  
16 IVR system (including e.g., the grammar used, specific questions to which it can  
17 respond, the interpretation of questions, and the answers to be given to customers)  
18 were derived by Defendant (and/or others on its behalf) from collecting and  
19 studying data from thousands of actual calls made to Defendant's customer support  
20 line. Based on this, Plaintiff believes that Defendant (and/or others on its behalf)  
21 has trained the IVR system with Defendant's call center data that is unique to  
22 Defendant's business. As a result, the IVR system is tailored to respond with  
23 appropriate answers to questions posed by Defendant's customer base.

24 19. Accordingly, Defendant's IVR system has been customized with  
25 customer content data that is not available from a third party. This Defendant-  
26 specific content data is critically important to the behavior and operation of  
27 Defendant's IVR system, since without it the IVR system would not know what

1 words to recognize from a caller's utterance, how to determine the meaning of such  
2 words, and/or what answer to give to the caller as a response.

3 20. Defendant's IVR system, as noted above, is a combination of  
4 components, including at least some hardware, software and content which it  
5 obtained from third parties (third party components). Nonetheless, and on  
6 information and belief, Defendant is responsible for and has caused such third party  
7 components to be combined, adapted and configured (including with such  
8 Defendant-specific content) in accordance with specific performance, content  
9 requirements and scenarios of the Defendant's customer support operations.

10 21. Consequently, and on further information and belief, the current  
11 structure and operation of Defendant's IVR system is a result of content  
12 contributions, performance specifications and operational specifications provided  
13 by Defendant and configuration/modification of third party components made by  
14 Defendant (and/or others on its behalf). Such third party components – as currently  
15 available from such third parties - by themselves would not be sufficient to  
16 implement Defendant's IVR system without Defendant's cooperation, contributions  
17 and actions, including Defendant's provision of the Defendant-specific content  
18 data.

19 22. On or about June 2, 2006, Plaintiff sent a letter to Defendant, stating  
20 that the IVR system is covered by one or more claims of the Patents in Suit. In that  
21 letter, Plaintiff included a number of supporting materials to explain its position on  
22 the Patents, and further extended an offer to license the Patents in Suit to  
23 Defendant. On or about June 27, 2006, Defendant responded, informing Plaintiff  
24 that it needed to investigate the matter and requested identification of the patent  
25 claims that may be infringed. On or about June 29, 2006, Plaintiff responded to  
26 Defendant, stating that Defendant may have overlooked the CD enclosed with the  
27 original letter which has extensive representative claim charts pointing out

1 particularly which claims Plaintiff believes are pertinent to Defendant's system and  
2 why. Some many months later on October 18, 2007, and having not heard from  
3 Defendant, Plaintiff sent another letter to Defendant to again negotiate a license and  
4 requested a response by no later than December 14, 2007. Defendant failed to  
5 respond in any meaningful way to the licensing offer or the charge of infringement,  
6 necessitating the filing of this action.

### 7 **III. JURISDICTION AND VENUE**

8 23. This Court has original subject matter jurisdiction over Plaintiff's  
9 patent infringement claim pursuant to 28 U.S.C. §1338(a).

10 24. This Court has personal jurisdiction over Defendant because  
11 Defendant's corporate headquarters are located in San Francisco, CA.

12 25. Venue properly lies in the Northern District of California pursuant to  
13 28 U.S.C. §1391 and §1400, because the acts complained of herein have been  
14 committed and are being committed in this Judicial District and Defendant is  
15 subject to personal jurisdiction within the District.

### 16 **IV. FIRST COUNT FOR INFRINGEMENT**

### 17 **OF UNITED STATES PATENT NO. 6,633,846**

18 26. Plaintiff hereby incorporates by reference the allegations contained in  
19 paragraphs 1 through 25.

20 27. Plaintiff is the assignee of the U.S. Patent No. 6,633,846 ("the '846  
21 Patent"), attached hereto as Exhibit 1, entitled "Distributed Real Time Speech  
22 Recognition System". Plaintiff owns and has standing and capacity to sue and  
23 recover damages for infringement under the '846 Patent.

24 28. Defendant has violated Plaintiff's patent rights by operating an IVR  
25 system covered by at least one claim of the '846 Patent. Wells Fargo's infringing  
26 IVR system has not been manufactured or authorized in any manner by the  
27 Plaintiff.

29. As a legal consequence of Defendant's infringement, Plaintiff is entitled to compensation for no less than a reasonable royalty, as well as pre-judgment interest and a preliminary and permanent injunction. In the event that the Court does not exercise its equitable discretion to award a permanent injunction, then Plaintiff is entitled to a judgment that includes a sum equal to the total projected value of a compulsory license for the life of the patent at a royalty rate to be determined by a jury, discounted to present value, to compensate Plaintiff for future infringement.

30. The infringement of the '846 Patent has been willful in that Defendant is fully aware of Plaintiff's rights, yet has continued to use the infringing IVR system in violation of the patent laws without a good faith basis for believing it does not infringe or the patent is invalid. This intentional refusal to respect Plaintiff's patent rights constitutes willful infringement under 35 U.S.C. §§ 284 and 285, thereby entitling Plaintiff to treble damages and attorneys' fees.

**V. SECOND COUNT FOR INFRINGEMENT OF**  
**UNITED STATES PATENT NO. 6,665,640**

31. Plaintiff hereby incorporates by reference the allegations contained in paragraphs 1 through 25.

32. Plaintiff is the assignee of the U.S. Patent No. 6,665,640 ("the '640 Patent"), attached hereto as Exhibit 2, entitled "Interactive Speech Based Learning/Training System Formulating Search Queries Based on Natural Language Parsing of Recognized User Queries". Plaintiff owns and has standing and capacity to sue and recover damages for infringement under the '640 Patent.

33. Defendant has violated Plaintiff's patent rights by operating an IVR system covered by at least one claim of the '640 Patent. Wells Fargo's infringing IVR system has not been manufactured or authorized in any manner by the Plaintiff.

34. As a legal consequence of Defendant's infringement, Plaintiff is entitled to compensation for no less than a reasonable royalty, as well as pre-judgment interest and a preliminary and permanent injunction. In the event that the Court does not exercise its equitable discretion to award a permanent injunction, then Plaintiff is entitled to a judgment that includes a sum equal to the total projected value of a compulsory license for the life of the patent at a royalty rate to be determined by a jury, discounted to present value, to compensate Plaintiff for future infringement.

35. The infringement of the '640 Patent has been willful in that Defendant is fully aware of Plaintiff's rights, yet has continued to use the infringing IVR system in violation of the patent laws without a good faith basis for believing it does not infringe or the patent is invalid. This intentional refusal to respect Plaintiff's patent rights constitutes willful infringement under 35 U.S.C. §§ 284 and 285, thereby entitling Plaintiff to treble damages and attorneys' fees.

**VI. THIRD COUNT FOR INFRINGEMENT**  
**OF UNITED STATES PATENT NO. 7,050,977**

36. Plaintiff hereby incorporates by reference the allegations contained in paragraphs 1 through 25.

37. Plaintiff is the assignee of the U.S. Patent No. 7,050,977 ("the '977 Patent"), attached hereto as Exhibit 3, entitled "Speech-Enabled Server for Internet Website and Method". Plaintiff owns and has standing and capacity to sue and recover damages for infringement under the '977 Patent.

38. Defendant has violated Plaintiff's patent rights by operating an IVR system covered by at least one claim of the '977 Patent. Wells Fargo's infringing IVR system has not been manufactured or authorized in any manner by the Plaintiff.



39. As a legal consequence of Defendant's infringement, Plaintiff is entitled to compensation for no less than a reasonable royalty, as well as pre-judgment interest and a preliminary and permanent injunction. In the event that the Court does not exercise its equitable discretion to award a permanent injunction, then Plaintiff is entitled to a judgment that includes a sum equal to the total projected value of a compulsory license for the life of the patent at a royalty rate to be determined by a jury, discounted to present value, to compensate Plaintiff for future infringement.

40. The infringement of the '977 Patent has been willful in that Defendant is fully aware of Plaintiff's rights, yet has continued to use the infringing IVR system in violation of the patent laws without a good faith basis for believing it does not infringe or the patent is invalid. This intentional refusal to respect Plaintiff's patent rights constitutes willful infringement under 35 U.S.C. §§ 284 and 285, thereby entitling Plaintiff to treble damages and attorneys' fees.

**VII. FOURTH COUNT FOR INFRINGEMENT**  
**OF UNITED STATES PATENT NO. 7,277,854**

41. Plaintiff hereby incorporates by reference the allegations contained in paragraphs 1 through 25.

42. Plaintiff is the assignee of the U.S. Patent No. 7,277,854 ("the '854 Patent"), attached hereto as Exhibit 4, entitled "Speech Recognition System Interactive Agent". Plaintiff owns and has standing and capacity to sue and recover damages for infringement under the '854 Patent.

43. Defendant has violated Plaintiff's patent rights by operating an IVR system covered by at least one claim of the '854 Patent. Wells Fargo's infringing IVR system has not been manufactured or authorized in any manner by the Plaintiff.



44. As a legal consequence of Defendant's infringement, Plaintiff is entitled to compensation for no less than a reasonable royalty, as well as pre-judgment interest and a preliminary and permanent injunction. In the event that the Court does not exercise its equitable discretion to award a permanent injunction, then Plaintiff is entitled to a judgment that includes a sum equal to the total projected value of a compulsory license for the life of the patent at a royalty rate to be determined by a jury, discounted to present value, to compensate Plaintiff for future infringement.

45. The infringement of the '854 Patent has been willful in that Defendant is fully aware of Plaintiff's rights, yet has continued to use the infringing IVR system in violation of the patent laws without a good faith basis for believing it does not infringe or the patent is invalid. This intentional refusal to respect Plaintiff's patent rights constitutes willful infringement under 35 U.S.C. §§ 284 and 285, thereby entitling Plaintiff to treble damages and attorneys' fees.

### **VIII. DEMAND FOR JURY TRIAL**

46. Plaintiff hereby exercises its right to a jury trial under the Seventh Amendment to the United States Constitution, and pursuant to Fed. R. Civ. Proc., Rule 38, demands a jury trial in accordance therewith.

### **IX. PRAYER FOR RELIEF**

WHEREFORE, Plaintiff prays for:

a. A preliminary injunction, barring Defendant and all of its agents, officers, attorneys, successors, and assigns from manufacturing, importing or using any system (or components thereof) that infringes upon the '846, the '640, the '977 and the '854 Patents;

b. A permanent injunction, barring Defendant and all of its agents, officers, successors and assigns from manufacturing, importing or using any system

(or components thereof) that infringes upon the '846, the '640, the '977 and the '854 Patents;

c. That Defendant be required to account to Plaintiff for all savings and revenues realized by Defendant and any subsidiary and any partner company of Defendant from the use of IVR systems infringing the '846, the '640, the '977 and the '854 Patents;

d. A judgment for compensatory damages, not less than reasonable royalty, suffered as a result of the patent infringement as well as prejudgment interest;

e. A judgment including a sum equal to a the total projected value of a compulsory license for the life of the patents, discounted to present value, to compensate Plaintiff for future infringement in the event that a permanent injunction is not awarded;

f. Treble damages and attorneys' fees pursuant to 35 U.S.C. §§ 284 and 285 for willful infringement of the '846, the '640, the '977 and the '854 Patents by Defendant; and,

g. Any and all other relief that the Court deems proper.

Respectfully submitted,

TROJAN LAW OFFICES

by

Dated: April 22, 2008

/s/R. Joseph Trojan

R. Joseph Trojan  
Attorney for Plaintiff,  
PHOENIX SOLUTIONS, INC.

# EXHIBIT 2

**FILE COPY**

1 KEKER & VAN NEST, LLP  
DARALYN J. DURIE - #169825  
2 EUGENE M. PAIGE - #202849  
RYAN M. KENT - #220441  
3 SANDEEP MITRA - #244054  
710 Sansome Street  
4 San Francisco, CA 94111-1704  
Telephone: (415) 391-5400  
5 Facsimile: (415) 397-7188

6 Attorneys for Defendant  
WELLS FARGO BANK, N.A.  
7

8 UNITED STATES DISTRICT COURT  
9  
10 NORTHERN DISTRICT OF CALIFORNIA  
11 SAN FRANCISCO DIVISION

12 PHOENIX SOLUTIONS, INC., a California  
13 corporation,

14 Plaintiff,

15 v.

16 WELLS FARGO BANK, N.A., a Delaware  
17 corporation,

18 Defendant.

Case No. CV 08-0863 MHP

**ANSWER TO AMENDED COMPLAINT  
DEMAND FOR JURY TRIAL**

19  
20  
21 Defendant Wells Fargo Bank, N.A. ("Wells Fargo") answers Phoenix Solutions, Inc.'s  
22 ("Phoenix's") amended complaint ("Complaint") as follows:

23 1. Wells Fargo admits that the Complaint purports to recite an action for  
24 infringement under the patent laws of the United States.

25 **I. THE PARTIES**

26 2. Wells Fargo denies that Phoenix is a corporation organized and existing under the  
27 laws of the State of California; Wells Fargo lacks knowledge or information sufficient to form a  
28

1 belief about the truth of the remainder of the allegations in this paragraph and, on that basis,  
2 denies the remainder of the allegations in this paragraph.

3 3. Wells Fargo admits that it has a place of business at 420 Montgomery Street, San  
4 Francisco, California 94163. Wells Fargo denies the remainder of the allegations of this  
5 paragraph.

6 **II. FACTUAL BACKGROUND**

7 4. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
8 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
9 paragraph.

10 5. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
11 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
12 paragraph.

13 6. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
14 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
15 paragraph.

16 7. Wells Fargo admits that it provides financial services including banking,  
17 insurance, investment, mortgage loan, and consumer finance services. Wells Fargo admits that it  
18 operates customer support lines, some of which are toll-free. Wells Fargo admits that some of its  
19 customer support lines employ interactive voice response (IVR) systems that provide customers  
20 with audible responses. Wells Fargo lacks information sufficient to form a belief about the truth  
21 of the remainder of the allegations in this paragraph and, on that basis, denies the remainder of  
22 the allegations in this paragraph.

23 8. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
24 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
25 paragraph.

26 9. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
27 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
28 paragraph.

1           10.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
2 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
3 paragraph.

4           11.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
5 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
6 paragraph.

7           12.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
8 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
9 paragraph.

10          13.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
11 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
12 paragraph.

13          14.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
14 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
15 paragraph.

16          15.     Wells Fargo admits that some of the IVR systems used in its customer support  
17 lines may respond with an audible response or may route the caller to a live person. Wells Fargo  
18 lacks information sufficient to form a belief about the truth of the remainder of the allegations in  
19 this paragraph and, on that basis, denies the remainder of the allegations in this paragraph.

20          16.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
21 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
22 paragraph.

23          17.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
24 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
25 paragraph.

26          18.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
27 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
28 paragraph.

1           19.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
2 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
3 paragraph.

4           20.     Wells Fargo admits that the IVR systems used in its customer support lines are a  
5 combination of components, including hardware, software, and content, that it obtained from  
6 third parties. Wells Fargo lacks knowledge or information sufficient to form a belief about the  
7 truth of the allegations in the remainder this paragraph and, on that basis, denies the allegations  
8 in the remainder of this paragraph.

9           21.     Wells Fargo lacks knowledge or information sufficient to form a belief about the  
10 truth of the allegations in this paragraph and, on that basis, denies the allegations in this  
11 paragraph.

12           22.     Wells Fargo admits that, on or about June 2, 2006, J. Nicholas Gross of the Trojan  
13 Law Offices sent a letter addressed to James Strother, purportedly on behalf of Phoenix, in which  
14 Mr. Gross stated that the "speech based electronic agent" that Mr. Gross apparently assumed was  
15 operated by Wells Fargo "is very likely covered one or more claims of the Phoenix portfolio in  
16 this area." Wells Fargo admits that the letter listed U.S. Patent Nos. 6,633,846, 6,616,172,  
17 6,665,640, and 7,050,977 and a pending publication, Publication No. 2004/0117189. Wells  
18 Fargo further admits that the letter stated that "we request that you please review the enclosed  
19 materials, and let us know within 30 days if Wells Fargo is interested in securing a license to the  
20 above technologies." Wells Fargo admits that, on or about June 27, 2006, Walter Linder pointed  
21 out in a letter to Mr. Gross that Mr. Gross had failed to identity any specific claims that were  
22 infringed and had not provided any specific reasons why any such claims were infringed. Wells  
23 Fargo admits that, on or about June 29, 2006, Mr. Gross replied by letter to Mr. Linder that  
24 Wells Fargo may have overlooked a CD enclosed with the original letter. Wells Fargo admits  
25 that, on or about October 18, 2007, R. Joseph Trojan, purportedly representing Phoenix, sent a  
26 letter to Mr. Linder stating, *inter alia*, "the only rational choice is for Wells Fargo to solicit more  
27 favorable treatment as a willing licensee than the terms it would receive as a defendant in  
28 litigation." The letter further demanded that Wells Fargo "disclose its call volume for each of

1 the past three years for its interactive natural language processing customer support lines.”

2 Wells Fargo denies the remainder of the allegations in this paragraph.

3 **III. JURISDICTION AND VENUE**

4 23. This paragraph states no more than a legal conclusion to which no response is  
5 required.

6 24. This paragraph states no more than a legal conclusion to which no response is  
7 required.

8 25. This paragraph states no more than a legal conclusion to which no response is  
9 required.

10 **IV. FIRST COUNT FOR INFRINGEMENT OF UNITED**  
11 **STATES PATENT NO. 6,633,846**

12 26. Wells Fargo repeats and realleges its responses set forth in paragraphs 1-25  
13 above.

14 27. Wells Fargo admits that what purports to be a copy of U.S. Patent No. 6,633,846  
15 (“ ’846 patent”) is attached to the Complaint as Exhibit 1. Wells Fargo admits that the ’846  
16 patent is entitled “Distributed Real Time Speech Recognition System.” Wells Fargo lacks  
17 knowledge or information sufficient to form a belief about the truth of the remainder of the  
18 allegations in this paragraph and, on that basis, denies the remainder of the allegations in this  
19 paragraph.

20 28. Denied.

21 29. Denied.

22 30. Denied.

23 **V. SECOND COUNT FOR INFRINGEMENT OF UNITED**  
24 **STATES PATENT NO. 6,665,640**

25 31. Wells Fargo repeats and realleges its responses set forth in paragraphs 1-25  
26 above.

27 32. Wells Fargo admits that what purports to be a copy of U.S. Patent No. 6,665,640  
28 (“ ’640 patent”) is attached to the Complaint as Exhibit 2. Wells Fargo admits that the ’640  
patent is entitled “Interactive Speech Based Learning/Training System Formulating Search



1 Queries Based on Natural Language Parsing of Recognized User Queries.” Wells Fargo lacks  
2 knowledge or information sufficient to form a belief about the truth of the remainder of the  
3 allegations in this paragraph and, on that basis, denies the remainder of the allegations in this  
4 paragraph.

5 33. Denied.

6 34. Denied.

7 35. Denied.

8 **VI. THIRD COUNT FOR INFRINGEMENT OF UNITED**  
9 **STATES PATENT NO. 7,050,977**

10 36. Wells Fargo repeats and realleges its responses set forth in paragraphs 1 - 25  
11 above.

12 37. Wells Fargo admits that what purports to be a copy of U.S. Patent No. 7,050,977  
13 (“ ’977 patent”) is attached to the Complaint as Exhibit 3. Wells Fargo admits that the ’977  
14 patent is entitled “Speech-Enabled Server for Internet Website and Method.” Wells Fargo lacks  
15 knowledge or information sufficient to form a belief about the truth of the remainder of the  
16 allegations in this paragraph and, on that basis, denies the remainder of the allegations in this  
17 paragraph.

18 38. Denied.

19 39. Denied.

20 40. Denied.

21 **VII. FOURTH COUNT FOR INFRINGEMENT OF UNITED**  
22 **STATES PATENT NO. 7,277,854**

23 41. Wells Fargo repeats and realleges its responses set forth in paragraphs 1 - 25  
24 above.

25 42. Wells Fargo admits that what purports to be a copy of U.S. Patent No. 7,277,854  
26 (“ ’854 patent”) is attached to the Complaint as Exhibit 4. Wells Fargo admits that the ’854  
27 patent is entitled “Speech Recognition System Interactive Agent.” Wells Fargo lacks knowledge  
28 or information sufficient to form a belief about the truth of the remainder of the allegations in  
this paragraph and, on that basis, denies the remainder of the allegations in this paragraph.

1 43. Denied.

2 44. Denied.

3 45. Denied.

4 **VIII. DEMAND FOR JURY TRIAL**

5 46. This paragraph demands a jury trial, and accordingly no response is necessary for  
6 this paragraph.

7 **IX. PRAYER FOR RELIEF**

8 47. Wells Fargo denies each allegation of the Complaint not expressly admitted  
9 herein.

10 **AFFIRMATIVE DEFENSES**

11 **FIRST AFFIRMATIVE DEFENSE**

12 48. On information and belief, the '846 patent is invalid because it fails to enable a  
13 person of ordinary skill in the art to make and/or use the purported inventions claimed therein as  
14 required by 35 U.S.C. § 112.

15 **SECOND AFFIRMATIVE DEFENSE**

16 49. On information and belief, the '846 patent is invalid because it fails to set forth an  
17 adequate written description of the purported inventions claimed therein as required by 35 U.S.C.  
18 § 112.

19 **THIRD AFFIRMATIVE DEFENSE**

20 50. On information and belief, the '846 patent is invalid because it fails to provide the  
21 best mode known to the putative inventors of practicing the purported inventions claimed therein  
22 as required by 35 U.S.C. § 112.

23 **FOURTH AFFIRMATIVE DEFENSE**

24 51. On information and belief, the '846 patent is invalid because it fails to satisfy the  
25 definiteness requirement of 35 U.S.C. § 112.

26 **FIFTH AFFIRMATIVE DEFENSE**

27 52. On information and belief, the '846 patent is invalid because the purported  
28 inventions claimed therein are anticipated by prior art under 35 U.S.C. § 102.

**SIXTH AFFIRMATIVE DEFENSE**

53. On information and belief, the '846 patent is invalid because the purported inventions claimed therein do not meet the requirement of non-obviousness contained in 35 U.S.C. § 103.

**SEVENTH AFFIRMATIVE DEFENSE**

54. On information and belief, the '846 patent is invalid because it fails to set forth the proper inventors of the purported inventions claimed in the patent.

**EIGHTH AFFIRMATIVE DEFENSE**

55. On information and belief, the '846 patent is not infringed by Wells Fargo because the claim constructions that would be required to find infringement are barred by the doctrine of prosecution disclaimer and/or prosecution history estoppel.

**NINTH AFFIRMATIVE DEFENSE**

56. On information and belief, the '640 patent is invalid because it fails to enable a person of ordinary skill in the art to make and/or use the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TENTH AFFIRMATIVE DEFENSE**

57. On information and belief, the '640 patent is invalid because it fails to set forth an adequate written description of the purported inventions claimed therein as required by 35 U.S.C. § 112.

**ELEVENTH AFFIRMATIVE DEFENSE**

58. On information and belief, the '640 patent is invalid because it fails to provide the best mode known to the putative inventors of practicing the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TWELFTH AFFIRMATIVE DEFENSE**

59. On information and belief, the '640 patent is invalid because it fails to satisfy the definiteness requirement of 35 U.S.C. § 112.

**THIRTEENTH AFFIRMATIVE DEFENSE**

60. On information and belief, the '640 patent is invalid because the purported inventions claimed therein are anticipated by prior art under 35 U.S.C. § 102.

**FOURTEENTH AFFIRMATIVE DEFENSE**

61. On information and belief, the '640 patent is invalid because the purported inventions claimed therein do not meet the requirement of non-obviousness contained in 35 U.S.C. § 103.

**FIFTEENTH AFFIRMATIVE DEFENSE**

62. On information and belief, the '640 patent is invalid because it fails to set forth the proper inventors of the purported inventions claimed in the patent.

**SIXTEENTH AFFIRMATIVE DEFENSE**

63. On information and belief, the '640 patent is not infringed by Wells Fargo because the claim constructions that would be required to find infringement are barred by the doctrine of prosecution disclaimer and/or prosecution history estoppel.

**SEVENTEENTH AFFIRMATIVE DEFENSE**

64. On information and belief, the '977 patent is invalid because it fails to enable a person of ordinary skill in the art to make and/or use the purported inventions claimed therein as required by 35 U.S.C. § 112.

**EIGHTEENTH AFFIRMATIVE DEFENSE**

65. On information and belief, the '977 patent is invalid because it fails to set forth an adequate written description of the purported inventions claimed therein as required by 35 U.S.C. § 112.

**NINETEENTH AFFIRMATIVE DEFENSE**

66. On information and belief, the '977 patent is invalid because it fails to provide the best mode known to the putative inventors of practicing the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TWENTIETH AFFIRMATIVE DEFENSE**

67. On information and belief, the '977 patent is invalid because it fails to satisfy the definiteness requirement of 35 U.S.C. § 112.

**TWENTY-FIRST AFFIRMATIVE DEFENSE**

68. On information and belief, the '977 patent is invalid because the purported inventions claimed therein are anticipated by prior art under 35 U.S.C. § 102.

**TWENTY-SECOND AFFIRMATIVE DEFENSE**

69. On information and belief, the '977 patent is invalid because the purported inventions claimed therein do not meet the requirement of non-obviousness contained in 35 U.S.C. § 103.

**TWENTY-THIRD AFFIRMATIVE DEFENSE**

70. On information and belief, the '977 patent is invalid because it fails to set forth the proper inventors of the purported inventions claimed in the patent.

**TWENTY-FOURTH AFFIRMATIVE DEFENSE**

71. On information and belief, the '977 patent is not infringed by Wells Fargo because the claim constructions that would be required to find infringement are barred by the doctrine of prosecution disclaimer and/or prosecution history estoppel.

**TWENTY-FIFTH AFFIRMATIVE DEFENSE**

72. On information and belief, the '854 patent is invalid because it fails to enable a person of ordinary skill in the art to make and/or use the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TWENTY-SIXTH AFFIRMATIVE DEFENSE**

73. On information and belief, the '854 patent is invalid because it fails to set forth an adequate written description of the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TWENTY-SEVENTH AFFIRMATIVE DEFENSE**

74. On information and belief, the '854 patent is invalid because it fails to provide the best mode known to the putative inventors of practicing the purported inventions claimed therein as required by 35 U.S.C. § 112.

**TWENTY-EIGHTH AFFIRMATIVE DEFENSE**

75. On information and belief, the '854 patent is invalid because it fails to satisfy the definiteness requirement of 35 U.S.C. § 112.

**TWENTY-NINTH AFFIRMATIVE DEFENSE**

76. On information and belief, the '854 patent is invalid because the purported inventions claimed therein are anticipated by prior art under 35 U.S.C. § 102.

**THIRTIETH AFFIRMATIVE DEFENSE**

77. On information and belief, the '854 patent is invalid because the purported inventions claimed therein do not meet the requirement of non-obviousness contained in 35 U.S.C. § 103.

**THIRTY-FIRST AFFIRMATIVE DEFENSE**

78. On information and belief, the '854 patent is invalid because it fails to set forth the proper inventors of the purported inventions claimed in the patent.

**THIRTY-SECOND AFFIRMATIVE DEFENSE**

79. On information and belief, the '854 patent is not infringed by Wells Fargo because the claim constructions that would be required to find infringement are barred by the doctrine of prosecution disclaimer and/or prosecution history estoppel.

**THIRTY-THIRD AFFIRMATIVE DEFENSE**

80. On information and belief, one or more of Phoenix's claims are barred by the doctrine of laches.

**THIRTY-FOURTH AFFIRMATIVE DEFENSE**

81. On information and belief, Phoenix's claims for damages are limited and/or barred by its failure to comply with the provisions of 35 U.S.C. § 287.

**THIRTY-FIFTH AFFIRMATIVE DEFENSE**

82. On information and belief, Phoenix's claims for infringement of the '846 patent are barred in whole or in part by its failure to comply with the duty of candor before the United States Patent and Trademark Office ("USPTO"). Phoenix misrepresented or omitted material information in prosecuting the '846 patent. The materiality of the information that was omitted is confirmed by the fact that, as explained further below, in each instance the reference in question was cited to Phoenix by a patent examiner overseeing the prosecution of a patent application seeking to claim related subject matter, and the reference was cited as a ground for rejecting the claims of that pending application. That demonstrates that a reasonable examiner would have likely considered the withheld information relevant in assessing the patentability of the claims here. Further, on information and belief, Phoenix withheld the information with the intent to deceive the USPTO. Phoenix's intent to deceive the USPTO can be inferred from the fact that it repeatedly failed to cite material prior art of which it was made aware during the course of prosecuting related applications. Illustrative examples of such failures to disclose material prior art of which Wells Fargo is currently aware are discussed below. As a result of at least these omissions, the '846 patent is unenforceable due to inequitable conduct.

83. During the time that the '846 patent was pending before the USPTO, Phoenix was aware of U.S. Patent No. 5,615,296 to Stanford. Phoenix became aware of the Stanford patent no later than May of 2002, when the Examiner in the '640 patent prosecution mailed an Office Action rejecting the claims of the '640 patent, based in part on obviousness over the Stanford patent.

84. As explained in paragraph 82 above, the Stanford patent's materiality is demonstrated by the fact that it was used to reject the claims of a patent application from the same family. The Stanford patent also discloses information that is unquestionably material to issues relating to the patentability of the claims of the '846 patent, including the issue of obviousness. For example, the '846 patent as issued claims a system "wherein said speech representative values are transmitted continuously during said speech utterances." The Stanford

1 patent, at column 4, lines 10-12 notes that it discloses a “technique of speaker-independent,  
2 continuous-speech phrases and bi-grams.”

3 85. Well over three months later, in September of 2002, Phoenix submitted a  
4 supplemental Information Disclosure Statement. That IDS contained no mention of the Stanford  
5 patent. Days after that, Phoenix submitted a set of amendments and arguments intended to  
6 overcome the Examiner’s prior rejection of the claims of the ’846 patent. Still no mention was  
7 made of the Stanford patent, despite the fact that Phoenix had attempted at length to distinguish  
8 the Stanford patent in the ’640 patent prosecution.

9 86. On March 12, 2003, the Examiner gave notice of allowance of all claims of the  
10 ’846 patent. Phoenix still failed to disclose to the USPTO the Stanford patent, a reference that  
11 may well have led the USPTO to withdraw its notice of allowance of the claims.

12 87. The ’846 patent reflects on its face that the Stanford patent was never considered  
13 by the Examiner during its prosecution. Notably, the attorney prosecuting both the ’846 patent  
14 and the ’640 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
15 material reference, Phoenix committed inequitable conduct, and the ’846 patent is unenforceable.

16 88. Also during the time that the ’846 patent was pending before the USPTO, Phoenix  
17 was aware of U.S. Patent No. 5,983,190 to Trower. Phoenix became aware of the Trower patent  
18 no later than May of 2002, when the Examiner in the ’640 patent prosecution mailed an Office  
19 Action rejecting the claims of the ’640 patent, based in part on obviousness over the Trower  
20 patent.

21 89. As explained in paragraph 82 above, the Trower patent’s materiality is  
22 demonstrated by the fact that it was used to reject the claims of a patent application from the  
23 same family. The Trower patent also discloses information that is unquestionably material to  
24 issues relating to the patentability of the claims of the ’846 patent, including the issue of  
25 obviousness. For example, the ’846 patent as issued claims a program used in a system “for  
26 receiving user speech utterance signals representing speech utterances to be recognized” that  
27 “works within a browser program executing on said computing system.” The Trower patent, at  
28 column 3, lines 15-16 and column 4, lines 28-34 notes that it discloses a system that relates to



1 “speech input” and utilizes “a microphone and analog to digital convertor circuitry for  
2 converting sound to digitized audio” and that the system is “advantageous for web pages.”

3 90. Well over three months later, in September of 2002, Phoenix submitted a  
4 supplemental Information Disclosure Statement. That IDS contained no mention of the Trower  
5 patent. Days after that, Phoenix submitted a set of amendments and arguments intended to  
6 overcome the Examiner’s prior rejection of the claims of the ’846 patent. Still no mention was  
7 made of the Trower patent.

8 91. On March 12, 2003, the Examiner gave notice of allowance of all claims of the  
9 ’846 patent. Phoenix still failed to disclose to the USPTO the Trower patent, a reference that  
10 may well have led the USPTO to withdraw its notice of allowance of the claims.

11 92. The ’846 patent reflects on its face that the Trower patent was never considered  
12 by the Examiner during its prosecution. Notably, the attorney prosecuting both the ’846 patent  
13 and the ’640 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
14 material reference, Phoenix committed inequitable conduct, and the ’846 patent is unenforceable.

15 **THIRTY-SIXTH AFFIRMATIVE DEFENSE**

16 93. On information and belief, Phoenix’s claims for infringement of the ’640 patent  
17 are barred in whole or in part by its failure to comply with the duty of candor before the USPTO.  
18 Phoenix misrepresented or omitted material information in prosecuting the ’640 patent. The  
19 materiality of the information that was omitted is confirmed by the fact that, as explained further  
20 below, in each instance the reference in question was cited to Phoenix by a patent examiner  
21 overseeing the prosecution of a patent application seeking to claim related subject matter, and the  
22 reference was cited as a ground for rejecting the claims of that pending application. That  
23 demonstrates that a reasonable examiner would have likely considered the withheld information  
24 relevant in assessing the patentability of the claims here. Further, on information and belief,  
25 Phoenix withheld the information with the intent to deceive the USPTO. Phoenix’s intent to  
26 deceive the USPTO can be inferred from the fact that it repeatedly failed to cite material prior art  
27 of which it was made aware during the course of prosecuting related applications. Illustrative  
28 examples of such failures to disclose material prior art of which Wells Fargo is currently aware

1 are discussed below. As a result of at least these omissions, the '640 patent is unenforceable due  
2 to inequitable conduct.

3 94. During the time that the '640 patent was pending before the USPTO, Phoenix was  
4 aware of U.S. Patent No. 5,737,485 to Flanagan. Phoenix became aware of the Flanagan patent  
5 no later than September of 2001, when the Examiner in the '846 patent prosecution mailed an  
6 Office Action rejecting the claims of the '846 patent, based in part on obviousness over the  
7 Flanagan patent.

8 95. As explained in paragraph 93 above, the Flanagan patent's materiality is  
9 demonstrated by the fact that it was used to reject the claims of a patent application from the  
10 same family. The Flanagan patent also discloses information that is unquestionably material to  
11 issues relating to the patentability of the claims of the '640 patent, including the issue of  
12 obviousness. For example, the '640 patent as issued claims "a speech recognition system for  
13 generating recognized speech utterance data from partially processed speech data." The  
14 Flanagan patent, at column 3, lines 55-57 and column 4, lines 2-4 discloses a "feature extractor  
15 [that] extracts speech features or cepstrum coefficients," which data are then "provided as inputs  
16 to the speech recognizer."

17 96. A year later, in September of 2002, Phoenix submitted a set of amendments and  
18 responses to the USPTO's Office Action rejecting the claims of the '640 patent. Phoenix made  
19 no mention of the Flanagan patent at that time. Shortly thereafter, Phoenix submitted another  
20 supplemental Information Disclosure Statement to the USPTO. Yet Phoenix again made no  
21 mention of the Flanagan patent.

22 97. The '640 patent reflects on its face that the Flanagan patent was never considered  
23 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '640 patent  
24 and the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
25 material reference, Phoenix committed inequitable conduct, and the '640 patent is unenforceable.

26 98. During the time that the '640 patent was pending before the USPTO, Phoenix was  
27 aware of U.S. Patent No. 5,265,014 to Haddock. Phoenix became aware of the Haddock patent  
28 no later than September of 2001, when the Examiner in the '846 patent prosecution mailed an

1 Office Action rejecting the claims of the '846 patent, based in part on obviousness over the  
2 Haddock patent.

3 99. As explained in paragraph 93 above, the Haddock patent's materiality is  
4 demonstrated by the fact that it was used to reject the claims of a patent application from the  
5 same family. The Haddock patent also discloses information information that is unquestionably  
6 material to issues relating to the patentability of the claims of the '640 patent, including the issue  
7 of obviousness. For example, the '640 patent as issued claims a system "adapted for responding  
8 to speech-based queries" that has a "speech recognition system for generating recognized speech  
9 utterance data" and "a query formulation system for converting said recognized speech data into  
10 a search query suitable for identifying a topic query entry corresponding to said speech-based  
11 query." The Haddock patent, at column 4, lines 25-28 and 43-46 notes that it discloses a system  
12 whereby "the user communicates textual information to the computer system by talking to the  
13 computer rather than by typing the information at the keyboard" and is "embodied in a user  
14 interface of a database system which receives a database query from a user, evaluates the query,  
15 and provides a result of the evaluation to the user."

16 100. A year later, in September of 2002, Phoenix submitted a set of amendments and  
17 responses to the USPTO's Office Action rejecting the claims of the '640 patent. Phoenix made  
18 no mention of the Haddock patent at that time. Shortly thereafter, Phoenix submitted another  
19 supplemental Information Disclosure Statement to the USPTO. Yet Phoenix again made no  
20 mention of the Haddock patent.

21 101. The '640 patent reflects on its face that the Haddock patent was never considered  
22 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '640 patent  
23 and the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
24 material reference, Phoenix committed inequitable conduct, and the '640 patent is unenforceable.

25 102. During the time that the '640 patent was pending before the USPTO, Phoenix was  
26 aware of U.S. Patent No. 6,336,090 to Chou. Phoenix became aware of the Chou patent no later  
27 than May of 2002, when the Examiner in the '846 patent prosecution mailed an Office Action  
28 rejecting the claims of the '846 patent, based in part on obviousness over the Chou patent.

1           103. As explained in paragraph 93 above, the Chou patent's materiality is  
2 demonstrated by the fact that it was used to reject the claims of a patent application from the  
3 same family. The Chou patent also discloses information information that is unquestionably  
4 material to issues relating to the patentability of the claims of the '640 patent, including the issue  
5 of obviousness. For example, the '640 patent as issued claims a system that involves "partially  
6 processed speech data being received from a remote speech capturing system." The Chou patent,  
7 at column 9, lines 51-59 notes that it discloses a "feature extraction and/or ASR units can be  
8 located a the receiving base station, the switch connected to the base station . . . or at another  
9 location connection on the network(s) to which these elements are connected" and that it will  
10 sometimes "be convenient to have the feature extraction and ASR operations performed at  
11 different locations."

12           104. A few months later, in September of 2002, Phoenix submitted a set of  
13 amendments and responses to the USPTO's Office Action rejecting the claims of the '640 patent.  
14 Phoenix made no mention of the Chou patent at that time. Shortly thereafter, Phoenix submitted  
15 another supplemental Information Disclosure Statement to the USPTO. Yet Phoenix again made  
16 no mention of the Chou patent.

17           105. The '640 patent reflects on its face that the Chou patent was never considered by  
18 the Examiner during its prosecution. Notably, the attorney prosecuting both the '640 patent and  
19 the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this material  
20 reference, Phoenix committed inequitable conduct, and the '640 patent is unenforceable.

21                           **THIRTY-SEVENTH AFFIRMATIVE DEFENSE**

22           106. On information and belief, Phoenix's claims for infringement of the '977 patent  
23 are barred in whole or in part by its failure to comply with the duty of candor before the USPTO.  
24 Phoenix misrepresented or omitted material information in prosecuting the '977 patent. The  
25 materiality of the information that was omitted is confirmed by the fact that, as explained further  
26 below, in each instance the reference in question was cited to Phoenix by a patent examiner  
27 overseeing the prosecution of a patent application seeking to claim related subject matter, and the  
28 reference was cited as a ground for rejecting the claims of that pending application. That

1 demonstrates that a reasonable examiner would have likely considered the withheld information  
2 relevant in assessing the patentability of the claims here. Further, on information and belief,  
3 Phoenix withheld the information with the intent to deceive the USPTO. Phoenix's intent to  
4 deceive the USPTO can be inferred from the fact that it repeatedly failed to cite material prior art  
5 of which it was made aware during the course of prosecuting related applications. Illustrative  
6 examples of such failures to disclose material prior art of which Wells Fargo is currently aware  
7 are discussed below. As a result of at least these omissions, the '977 patent is unenforceable due  
8 to inequitable conduct.

9 107. During the time that the '977 patent was pending before the USPTO, Phoenix was  
10 aware of U.S. Patent No. 5,615,296 to Stanford. Phoenix became aware of the Stanford patent  
11 no later than May of 2002, when the Examiner in the '640 patent prosecution mailed an Office  
12 Action rejecting the claims of the '640 patent, based in part on obviousness over the Stanford  
13 patent.

14 108. As explained in paragraph 106 above, the Stanford patent's materiality is  
15 demonstrated by the fact that it was used to reject the claims of a patent application from the  
16 same family. The Stanford patent also discloses information that is unquestionably material to  
17 issues relating to the patentability of the claims of the '977 patent, including the issue of  
18 obviousness. For example, the '977 patent as issued claims a system "adapted to interact on a  
19 real-time basis in response to one or more continuous speech queries." The Stanford patent, at  
20 column 4, lines 10-12 notes that it discloses a "technique of speaker-independent, continuous-  
21 speech phrases and bi-grams."

22 109. After May of 2002, Phoenix submitted no less than five Information Disclosure  
23 Statements. Not one disclosed the Stanford patent. Phoenix also twice amended its claims, but  
24 did not make any mention of the Stanford patent when doing so, despite the fact that Phoenix had  
25 attempted at length to distinguish the Stanford patent in the '640 patent prosecution.

26 110. The '977 patent reflects on its face that the Stanford patent was never considered  
27 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent  
28

1 and the '640 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
2 material reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

3 111. During the time that the '977 patent was pending before the USPTO, Phoenix was  
4 aware of U.S. Patent No. 5,737,485 to Flanagan. Phoenix became aware of the Flanagan patent  
5 no later than September of 2001, when the Examiner in the '846 patent prosecution mailed an  
6 Office Action rejecting the claims of the '846 patent, based in part on obviousness over the  
7 Flanagan patent.

8 112. As explained in paragraph 106 above, the Flanagan patent's materiality is  
9 demonstrated by the fact that it was used to reject the claims of a patent application from the  
10 same family. The Flanagan patent also discloses information that is unquestionably material to  
11 issues relating to the patentability of the claims of the '977 patent, including the issue of  
12 obviousness. For example, the '977 patent as issued claims "partially processing a speech  
13 utterance at the client platform to generate limited data content speech data." The Flanagan  
14 patent, at column 3, lines 55-57 and column 4, lines 2-4 discloses a "feature extractor [that]  
15 extracts speech features or cepstrum coefficients," which partially processed speech data are then  
16 "provided as inputs to the speech recognizer."

17 113. After September of 2001, Phoenix submitted a half-dozen Information Disclosure  
18 Statements. Not one disclosed the Flanagan patent. Phoenix also twice amended its claims, but  
19 did not make any mention of the Flanagan patent when doing so.

20 114. The '977 patent reflects on its face that the Flanagan patent was never considered  
21 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent  
22 and the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
23 material reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

24 115. During the time that the '977 patent was pending before the USPTO, Phoenix was  
25 aware of U.S. Patent No. 5,265,014 to Haddock. Phoenix became aware of the Haddock patent  
26 no later than September of 2001, when the Examiner in the '846 patent prosecution mailed an  
27 Office Action rejecting the claims of the '846 patent, based in part on obviousness over the  
28 Haddock patent.

1           116. As explained in paragraph 106 above, the Haddock patent's materiality is  
2 demonstrated by the fact that it was used to reject the claims of a patent application from the  
3 same family. The Haddock patent also discloses information that is unquestionably material to  
4 issues relating to the patentability of the claims of the '977 patent, including the issue of  
5 obviousness. For example, the '977 patent as issued claims a website that has a "speech  
6 recognition routine executing on the server computing system for completing recognition of said  
7 speech query using said speech data and said data content to generate a recognized speech  
8 query" and "a list of items, at least some of said list of items being selectable by a user based on  
9 said recognized speech query." The Haddock patent, at column 4, lines 25-28 and 43-46 notes  
10 that it discloses a system whereby "the user communicates textual information to the computer  
11 system by talking to the computer rather than by typing the information at the keyboard" and is  
12 "embodied in a user interface of a database system which receives a database query from a user,  
13 evaluates the query, and provides a result of the evaluation to the user."

14           117. After September of 2001, Phoenix submitted a half-dozen Information Disclosure  
15 Statements. Not one disclosed the Haddock patent. Phoenix also twice amended its claims, but  
16 did not make any mention of the Haddock patent when doing so.

17           118. The '977 patent reflects on its face that the Haddock patent was never considered  
18 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent  
19 and the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
20 material reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

21           119. During the time that the '977 patent was pending before the USPTO, Phoenix was  
22 aware of U.S. Patent No. 5,540,589 to Waters. Phoenix became aware of the Waters patent no  
23 later than September of 2001, when the Examiner in the '846 patent prosecution mailed an  
24 Office Action rejecting the claims of the '846 patent, based in part on obviousness over the  
25 Waters patent.

26           120. As explained in paragraph 106 above, the Waters patent's materiality is  
27 demonstrated by the fact that it was used to reject the claims of a patent application from the  
28 same family. The Waters patent also discloses information that is unquestionably material to



1 issues relating to the patentability of the claims of the '977 patent, including the issue of  
2 obviousness. For example, the '977 patent as issued claims a system "wherein signal processing  
3 functions required to generate said recognized speech query can be allocated between a client  
4 platform and the server computing system as needed based on computing resources available to  
5 said client platform and server computing system respectively." The Waters patent, at column 6,  
6 lines 21-23 notes that it discloses a system where the "voice recognizer 34 is illustrated as a  
7 standalone component, although it may be built-in to the controller."

8 121. After September of 2001, Phoenix submitted a half-dozen Information Disclosure  
9 Statements. Not one disclosed the Waters patent. Phoenix also twice amended its claims, but  
10 did not make any mention of the Waters patent when doing so.

11 122. The '977 patent reflects on its face that the Waters patent was never considered by  
12 the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent and  
13 the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this material  
14 reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

15 123. During the time that the '977 patent was pending before the USPTO, Phoenix was  
16 aware of U.S. Patent No. 6,336,090 to Chou. Phoenix became aware of the Chou patent no later  
17 than May of 2002, when the Examiner in the '846 patent prosecution mailed an Office Action  
18 rejecting the claims of the '846 patent, based in part on obviousness over the Chou patent.

19 124. As explained in paragraph 106 above, the Chou patent's materiality is  
20 demonstrated by the fact that it was used to reject the claims of a patent application from the  
21 same family. The Chou patent also discloses information that is unquestionably material to  
22 issues relating to the patentability of the claims of the '977 patent, including the issue of  
23 obviousness. For example, the '977 patent as issued claims a website that allows certain speech-  
24 recognition operations to "be allocated between a client platform and the server computing  
25 system as needed based on computing resources available to said client platform and server  
26 computing system respectively." The Chou patent, at column 9, lines 51-59 notes that it  
27 discloses a "feature extraction and/or ASR units can be located a the receiving base station, the  
28 switch connected to the base station . . . or at another location connection on the network(s) to



1 which these elements are connected” and that it will sometimes “be convenient to have the  
2 feature extraction and ASR operations performed at different locations.”

3 125. After May of 2002, Phoenix submitted no less than five Information Disclosure  
4 Statements. Not one disclosed the Chou patent. Phoenix also twice amended its claims, but did  
5 not make any mention of the Chou patent when doing so.

6 126. The '977 patent reflects on its face that the Chou patent was never considered by  
7 the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent and  
8 the '846 patent was the same: J. Nicholas Gross. By intentionally failing to submit this material  
9 reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

10 127. During the time that the '977 patent was pending before the USPTO, Phoenix was  
11 aware of U.S. Patent No. 5,983,190 to Trower. Phoenix became aware of the Trower patent no  
12 later than May of 2002, when the Examiner in the '640 patent prosecution mailed an Office  
13 Action rejecting the claims of the '640 patent, based in part on obviousness over the Trower  
14 patent.

15 128. As explained in paragraph 106 above, the Trower patent's materiality is  
16 demonstrated by the fact that it was used to reject the claims of a patent application from the  
17 same family. The Trower patent also discloses information that is unquestionably material to  
18 issues relating to the patentability of the claims of the '977 patent, including the issue of  
19 obviousness. For example, the '977 patent as issued claims a website that “controls an  
20 interactive character agent presented to the user for assisting in handling said speech query.”  
21 The Trower patent, at column 2, lines 23-25 and column 3, lines 15-17 notes that it discloses a  
22 “client-server animation system used to display interactive, animated user interface characters  
23 with speech input and output capability” and that the invention is “advantageous for web pages  
24 because a web page can include an interactive character simply by adding a reference to the  
25 agent server.”

26 129. After May of 2002, Phoenix submitted no less than five Information Disclosure  
27 Statements. Not one disclosed the Trower patent. Phoenix also twice amended its claims, but  
28 did not make any mention of the Trower patent when doing so.

1           130. The '977 patent reflects on its face that the Trower patent was never considered  
2 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '977 patent  
3 and the '640 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
4 material reference, Phoenix committed inequitable conduct, and the '977 patent is unenforceable.

5                           **THIRTY-EIGHTH AFFIRMATIVE DEFENSE**

6           131. On information and belief, Phoenix's claims for infringement of the '854 patent  
7 are barred in whole or in part by its failure to comply with the duty of candor before the USPTO.  
8 Phoenix misrepresented or omitted material information in prosecuting the '854 patent. The  
9 materiality of the information that was omitted is confirmed by the fact that, as explained further  
10 below, in each instance the reference in question was cited to Phoenix by a patent examiner  
11 overseeing the prosecution of a patent application seeking to claim related subject matter, and the  
12 reference was cited as a ground for rejecting the claims of that pending application. That  
13 demonstrates that a reasonable examiner would have likely considered the withheld information  
14 relevant in assessing the patentability of the claims here. Further, on information and belief,  
15 Phoenix withheld the information with the intent to deceive the USPTO. Phoenix's intent to  
16 deceive the USPTO can be inferred from the fact that it repeatedly failed to cite material prior art  
17 of which it was made aware during the course of prosecuting related applications. Illustrative  
18 examples of such failures to disclose material prior art of which Wells Fargo is currently aware  
19 are discussed below. As a result of at least these omissions, the '854 patent is unenforceable due  
20 to inequitable conduct.

21           132. During the time that the '854 patent was pending before the USPTO, Phoenix was  
22 aware of U.S. Patent No. 5,983,190 to Trower. Phoenix became aware of the Trower patent no  
23 later than May of 2002, when the Examiner in the '640 patent prosecution mailed an Office  
24 Action rejecting the claims of the '640 patent, based in part on obviousness over the Trower  
25 patent.

26           133. As explained in paragraph 131 above, the Trower patent's materiality is  
27 demonstrated by the fact that it was used to reject the claims of a patent application from the  
28 same family. The Trower patent also discloses information that is unquestionably material to

1 issues relating to the patentability of the claims of the '854 patent, including the issue of  
2 obviousness. For example, the '854 patent as issued claims a method employing an "interactive  
3 electronic agent" that "is an animated character on a screen of the client device." The Trower  
4 patent, at column 2, lines 23-25 notes that it discloses a "client-server animation system used to  
5 display interactive, animated user interface characters with speech input and output capability."

6 134. Phoenix filed the continuation application that matured into the '854 patent in  
7 January of 2005, nearly three years after it indisputably learned of the Trower patent. At no time  
8 during the prosecution of the '854 patent did Phoenix disclose the Trower patent to the USPTO.

9 135. The '854 patent reflects on its face that the Trower patent was never considered  
10 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '854 patent  
11 and the '640 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
12 material reference, Phoenix committed inequitable conduct, and the '854 patent is unenforceable.

13 136. During the time that the '854 patent was pending before the USPTO, Phoenix was  
14 aware of U.S. Patent No. 6,101,472 to Giangarra. Phoenix became aware of the Giangarra patent  
15 no later than August of 2004, when the Examiner in the '977 patent prosecution mailed an Office  
16 Action rejecting the claims of the '977 patent, based in part on obviousness over the Giangarra  
17 patent.

18 137. As explained in paragraph 131 above, the Giangarra patent's materiality is  
19 demonstrated by the fact that it was used to reject the claims of a patent application from the  
20 same family. The Giangarra patent also discloses information that is unquestionably material to  
21 issues relating to the patentability of the claims of the '854 patent, including the issue of  
22 obviousness. For example, the '854 patent as issued claims a method that includes "providing a  
23 speech recognition engine adapted to recognize a first set of words and/or phrases during an  
24 interactive speech session." The Giangarra patent, at column 5, lines 41-44 discloses a  
25 "vocabulary list stored in speech recognition unit 252 [that] provides a list of all words and  
26 utterances by an external user which will be recognized as voice commands."

27 138. Phoenix filed the continuation application that matured into the '854 patent in  
28 January of 2005, several months after it indisputably learned of the Giangarra patent. At no time

1 during the prosecution of the '854 patent did Phoenix disclose the Giangarra patent to the  
2 USPTO.

3 139. The '854 patent reflects on its face that the Giangarra patent was never considered  
4 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '854 patent  
5 and the '977 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
6 material reference, Phoenix committed inequitable conduct, and the '854 patent is unenforceable.

7 140. During the time that the '854 patent was pending before the USPTO, Phoenix was  
8 aware of U.S. Patent No. 6,330,530 to Horiguchi. Phoenix became aware of the Horiguchi  
9 patent no later than August of 2004, when the Examiner in the '977 patent prosecution mailed an  
10 Office Action rejecting the claims of the '977 patent, based in part on obviousness over the  
11 Horiguchi patent.

12 141. As explained in paragraph 131 above, the Horiguchi patent's materiality is  
13 demonstrated by the fact that it was used to reject the claims of a patent application from the  
14 same family. The Horiguchi patent also discloses information that is unquestionably material to  
15 issues relating to the patentability of the claims of the '854 patent, including the issue of  
16 obviousness. For example, the '854 patent as issued claims "a natural language query system."  
17 The Horiguchi patent, at column 1, lines 27-28 describes a "natural language processing system."

18 142. Phoenix filed the continuation application that matured into the '854 patent in  
19 January of 2005, several months after it indisputably learned of the Horiguchi patent. At no time  
20 during the prosecution of the '854 patent did Phoenix disclose the Horiguchi patent to the  
21 USPTO.

22 143. The '854 patent reflects on its face that the Horiguchi patent was never considered  
23 by the Examiner during its prosecution. Notably, the attorney prosecuting both the '854 patent  
24 and the '977 patent was the same: J. Nicholas Gross. By intentionally failing to submit this  
25 material reference, Phoenix committed inequitable conduct, and the '854 patent is unenforceable.

26 144. During the time that the '854 patent was pending before the USPTO, Phoenix was  
27 aware of U.S. Patent No. 6,901,366 to Kuhn. Phoenix became aware of the Kuhn patent no later  
28

1 than June of 2005, when the Examiner in the '977 patent prosecution mailed an Office Action  
2 rejecting the claims of the '977 patent, based in part on obviousness over the Kuhn patent.

3 145. As explained in paragraph 131 above, the Kuhn patent's materiality is  
4 demonstrated by the fact that it was used to reject the claims of a patent application from the  
5 same family. The Kuhn patent also discloses information that is unquestionably material to  
6 issues relating to the patentability of the claims of the '854 patent, including the issue of  
7 obviousness. For example, the '854 patent as issued claims a method of using a system that  
8 provides "a database of query/answer pairs concerning one or more topics which can be  
9 responded to by the natural language query system." The Kuhn patent, at column 5, line 1 and  
10 lines 45-47 notes that it discloses a "knowledge database" as well as a "natural language parser  
11 12 [that] analyzes and extracts semantically important and meaningful topics from a loosely  
12 structured, natural language text."

13 146. After June of 2005, Phoenix submitted several Information Disclosure  
14 Statements, and also amended the claims several times. At no time during the prosecution of the  
15 '854 patent did Phoenix disclose the Kuhn patent to the USPTO.

16 147. The '854 patent reflects on its face that the Kuhn patent was never considered by  
17 the Examiner during its prosecution. Notably, the attorney prosecuting both the '854 patent and  
18 the '977 patent was the same: J. Nicholas Gross. By intentionally failing to submit this material  
19 reference, Phoenix committed inequitable conduct, and the '854 patent is unenforceable.

20 **THIRTY-NINTH AFFIRMATIVE DEFENSE**

21 148. On information and belief, the '846 patent is invalid under the doctrine barring  
22 double patenting and/or obviousness-type double patenting.

23 **FORTIETH AFFIRMATIVE DEFENSE**

24 149. On information and belief, the '640 patent is invalid under the doctrine barring  
25 double patenting and/or obviousness-type double patenting.

26 **FORTY-FIRST AFFIRMATIVE DEFENSE**

27 150. On information and belief, the '977 patent is invalid under the doctrine barring  
28 double patenting and/or obviousness-type double patenting.

**FORTY-SECOND AFFIRMATIVE DEFENSE**

151. On information and belief, the '854 patent is invalid under the doctrine barring double patenting and/or obviousness-type double patenting.

**PRAYER FOR RELIEF**

WHEREFORE, Wells Fargo prays for judgment as follows:

(a) That Phoenix take nothing by its Complaint and the Court dismiss its Complaint with prejudice;

(b) That the Court find that no claim of the '846 patent has been, or is, infringed willfully, deliberately, or otherwise by Wells Fargo;

(c) That the Court find that no claim of the '640 patent has been, or is, infringed willfully, deliberately, or otherwise by Wells Fargo;

(d) That the Court find that no claim of the '977 patent has been, or is, infringed willfully, deliberately, or otherwise by Wells Fargo;

(e) That the Court find that no claim of the '854 patent has been, or is, infringed willfully, deliberately, or otherwise by Wells Fargo;

(f) That the Court find that the claims of the '846 patent are invalid;

(g) That the Court find that the claims of the '640 patent are invalid;

(h) That the Court find that the claims of the '977 patent are invalid;

(i) That the Court find that the claims of the '854 patent are invalid;

(j) That the Court find that the '846 patent is unenforceable because of inequitable conduct committed during its prosecution;

(k) That the Court find that the '640 patent is unenforceable because of inequitable conduct committed during its prosecution;

(l) That the Court find that the '977 patent is unenforceable because of inequitable conduct committed during its prosecution;

(m) That the Court find that the '854 patent is unenforceable because of inequitable conduct committed during its prosecution;

(n) That the Court award Wells Fargo reasonable attorneys' fees under 35 U.S.C. § 285;

1 (o) That the Court award Wells Fargo all costs and expenses it incurs in this action;

2 (p) That the Court award Wells Fargo such other and further relief that it deems just and  
3 proper.

4 **DEMAND FOR JURY TRIAL**

5 Wells Fargo hereby demands a trial by jury of all issues so triable in this action.

6  
7 Dated: May 7, 2008

KEKER & VAN NEST, LLP

8  
9 By: /s/ Eugene M. Paige

10 Eugene M. Paige  
11 Attorneys for Defendant  
12 WELLS FARGO BANK, N.A.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

# EXHIBIT 3



# EXHIBIT 3



US007050977B1

(12) **United States Patent**  
**Bennett**

(10) **Patent No.:** **US 7,050,977 B1**

(45) **Date of Patent:** **May 23, 2006**

(54) **SPEECH-ENABLED SERVER FOR  
INTERNET WEBSITE AND METHOD**

(75) Inventor: **Ian M. Bennett**, Palo Alto, CA (US)

(73) Assignee: **Phoenix Solutions, Inc.**, Palo Alto, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/439,174**

(22) Filed: **Nov. 12, 1999**

(51) **Int. Cl.**  
**G10L 15/02** (2006.01)  
**G10L 15/18** (2006.01)  
**G10L 15/22** (2006.01)  
**G06F 17/20** (2006.01)

(52) **U.S. Cl.** ..... **704/270.1; 704/275; 707/3**

(58) **Field of Classification Search** ..... **704/231-235, 704/251, 255, 257, 260, 270, 270.1, 275; 707/3, 4, 5**

See application file for complete search history.

(56) **References Cited**

**U.S. PATENT DOCUMENTS**

4,785,408 A	11/1988	Britton et al.	
4,852,170 A	7/1989	Bordeaux	
5,371,901 A	12/1994	Reed et al.	
5,509,104 A	4/1996	Lee et al.	
5,553,119 A	9/1996	McAlliser et al.	
5,652,897 A	7/1997	Linebarger et al.	
5,668,854 A	9/1997	Minakami et al.	
5,675,707 A	10/1997	Gorin et al.	
5,680,511 A	10/1997	Baker et al.	
5,758,322 A	5/1998	Rongley	704/275
5,802,256 A	9/1998	Fawcett et al.	
5,819,220 A	10/1998	Sarukkai et al.	704/243
5,836,771 A	11/1998	Ho et al.	
5,860,063 A	1/1999	Gorin et al.	
5,867,817 A	2/1999	Catallo et al.	704/255
5,873,062 A	2/1999	Hansen et al.	

5,884,302 A	3/1999	Ho	
5,915,236 A	6/1999	Gould et al.	704/251
5,934,910 A	8/1999	Ho et al.	
5,956,683 A	9/1999	Jacobs et al.	704/275
5,960,394 A	9/1999	Gould et al.	704/240
5,960,399 A	9/1999	Barclay et al.	704/270

(Continued)

**FOREIGN PATENT DOCUMENTS**

EP 1 094 388 A2 4/2001

(Continued)

**OTHER PUBLICATIONS**

Arons, B., "The Design of Audio Servers and Toolkits for Supporting Speech in the User Interface," believed to be published in: Journal of the American Voice I/O Society, pp. 27-41, Mar. 1991.

(Continued)

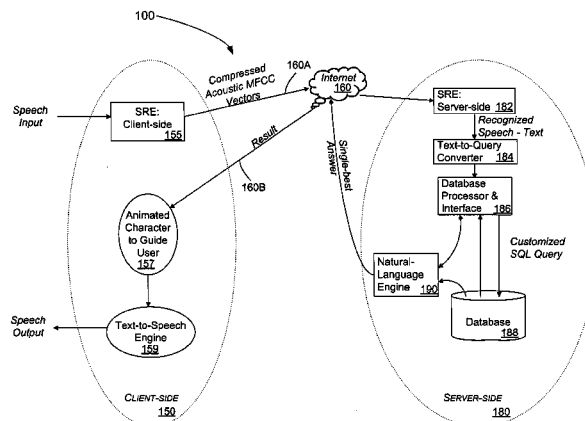
*Primary Examiner*—Martin Lerner

(74) *Attorney, Agent, or Firm*—J. Nicholas Gross

(57) **ABSTRACT**

An Internet-based server with speech support for enhanced interactivity is disclosed. This server hosts a server-side speech recognition engine and additional linguistic and database functions that cooperate to provide enhanced interactivity for clients so that their browsing experience is more satisfying, efficient and productive. This human-like interactivity which allows the user to ask queries about topics that range from customer delivery, product descriptions, payment details, is facilitated by the allowing the user to articulate the his or her questions directly in his or her natural language. The answer typically provided in real-time, can also be interfaced and integrated with existing telephone, e-mail and other mixed media services to provide a single point of interactivity for the user when browsing at a web-site.

**78 Claims, 31 Drawing Sheets**



## US 7,050,977 B1

Page 2

## U.S. PATENT DOCUMENTS

5,978,756 A 11/1999 Walker et al.  
 5,987,410 A 11/1999 Kellner et al.  
 5,995,918 A 11/1999 Kendall et al.  
 5,995,928 A 11/1999 Nguyen et al.  
 6,009,387 A 12/1999 Ramaswamy et al.  
 6,029,124 A 2/2000 Gillick et al. .... 704/200  
 6,032,111 A 2/2000 Mohri  
 6,035,275 A 3/2000 Brode et al.  
 6,044,266 A 3/2000 Kato  
 6,044,337 A 3/2000 Gorin et al.  
 6,101,472 A \* 8/2000 Giangarra et al. .... 704/275  
 6,112,176 A 8/2000 Goldenthal et al.  
 6,119,087 A 9/2000 Kuhn et al.  
 6,125,284 A 9/2000 Moore et al.  
 6,125,341 A 9/2000 Raud et al.  
 6,138,089 A 10/2000 Guberman ..... 704/207  
 6,141,640 A 10/2000 Moo  
 6,144,848 A 11/2000 Walsh et al.  
 6,144,938 A 11/2000 Surace et al. .... 704/257  
 6,178,404 B1 1/2001 Hambleton et al.  
 6,182,038 B1 1/2001 Balakrishnan et al.  
 6,182,068 B1 1/2001 Culliss  
 6,185,535 B1 2/2001 Hedin et al.  
 6,192,110 B1 2/2001 Abella et al.  
 6,195,636 B1 2/2001 Crupi et al.  
 6,226,610 B1 5/2001 Keiller et al.  
 6,233,559 B1 5/2001 Balakrishnan  
 6,243,679 B1 6/2001 Mohri et al.  
 6,246,986 B1 6/2001 Ammicht et al.  
 6,246,989 B1 6/2001 Polcyn  
 6,256,607 B1 7/2001 Digalakis et al.  
 6,269,336 B1 7/2001 Ladd et al.  
 6,278,973 B1 8/2001 Chung et al.  
 6,292,767 B1 9/2001 Jackson et al.  
 6,292,781 B1 9/2001 Urs et al.  
 6,327,561 B1 12/2001 Smith et al.  
 6,327,568 B1 12/2001 Joost  
 6,330,530 B1 \* 12/2001 Horiguchi et al. .... 704/4  
 6,363,349 B1 3/2002 Urs et al.  
 6,374,219 B1 4/2002 Jiang  
 6,374,226 B1 4/2002 Hunt et al.  
 6,381,594 B1 4/2002 Eichstaedt et al.  
 6,389,389 B1 5/2002 Meunier et al.  
 6,408,272 B1 6/2002 White et al.  
 6,411,926 B1 6/2002 Chang  
 6,418,199 B1 \* 7/2002 Perrone ..... 379/88.01  
 6,427,063 B1 7/2002 Cook et al.  
 6,434,529 B1 8/2002 Walker et al.  
 6,453,020 B1 9/2002 Hughes et al.  
 6,499,011 B1 12/2002 Souvignier et al.  
 6,499,013 B1 12/2002 Weber  
 6,510,411 B1 1/2003 Norton et al.  
 6,513,037 B1 1/2003 Ruber et al.  
 6,522,725 B1 2/2003 Kato  
 6,532,444 B1 3/2003 Weber  
 6,539,359 B1 \* 3/2003 Ladd et al. .... 704/275  
 6,567,778 B1 5/2003 Chao Chang et al.  
 6,574,597 B1 6/2003 Mohri et al.  
 6,584,464 B1 6/2003 Warthen  
 6,594,269 B1 7/2003 Polcyn  
 6,594,348 B1 7/2003 Bjurstrom et al.  
 6,614,885 B1 9/2003 Polcyn  
 6,618,726 B1 \* 9/2003 Colbath et al. .... 707/6  
 6,633,846 B1 \* 10/2003 Bennett et al. .... 704/257  
 6,681,206 B1 1/2004 Gorin et al.  
 6,697,780 B1 2/2004 Beutnagel et al.  
 6,742,021 B1 \* 5/2004 Halverson et al. .... 709/218  
 6,823,308 B1 11/2004 Keiller et al.  
 6,871,179 B1 3/2005 Kist et al.  
 6,901,366 B1 \* 5/2005 Kuhn et al. .... 704/275  
 6,922,733 B1 \* 7/2005 Kuiken et al. .... 709/246

6,940,953 B1 \* 9/2005 Eberle et al. .... 379/88.13  
 6,941,273 B1 \* 9/2005 Loghmani et al. .... 705/26  
 6,961,954 B1 \* 11/2005 Maybury et al. .... 725/53  
 6,964,012 B1 \* 11/2005 Zirngibl et al. .... 715/513  
 6,965,864 B1 \* 11/2005 Thrift et al. .... 704/275  
 6,965,890 B1 \* 11/2005 Dey et al. .... 707/4  
 2001/0016813 A1 8/2001 Brown et al.  
 2001/0032083 A1 10/2001 Van Clevén  
 2001/0056346 A1 12/2001 Ueyama et al.  
 2002/0032566 A1 3/2002 Tzirkel-Hancock et al.  
 2002/0046023 A1 4/2002 Fujii et al.  
 2002/0059068 A1 5/2002 Rose et al.  
 2002/0059069 A1 5/2002 Hsu et al.  
 2002/0086269 A1 7/2002 Shpiro  
 2002/0087325 A1 7/2002 Lee et al.  
 2002/0087655 A1 7/2002 Bridgman et al.  
 2002/0091527 A1 7/2002 Shiau  
 2003/0191625 A1 10/2003 Gorin et al.  
 2005/0091056 A1 4/2005 Surace et al.  
 2005/0131704 A1 6/2005 Dragosh et al.

## FOREIGN PATENT DOCUMENTS

EP 1 096 471 A1 5/2001  
 WO WO 99/48011 A1 9/1999  
 WO WO 99/50830 10/1999  
 WO WO 00/14727 A1 3/2000  
 WO WO 00/17854 A1 3/2000  
 WO WO 00/20962 A2 4/2000  
 WO WO 00/21075 A1 4/2000  
 WO WO 00/21232 A2 4/2000  
 WO WO 00/22610 A1 4/2000  
 WO WO 00/30072 A2 5/2000  
 WO WO 00/30287 A1 5/2000  
 WO WO 00/68823 A2 11/2000  
 WO WO 01/16936 A1 3/2001  
 WO WO 01/18693 A2 3/2001  
 WO WO 01/26093 A1 4/2001  
 WO WO 01/78065 A1 10/2001  
 WO WO 01/95312 A1 12/2001  
 WO WO 02/03380 A1 1/2002

## OTHER PUBLICATIONS

Hazen, T et al., "Recent Improvements in an Approach to Segment-Based Automatic Language Identification," believed to be published in: Proceedings of the 1994 International Conference on Spoken Language Processing, Yokohama, Japan, pp. 1883-1886, Sep. 1994.  
 House, D., "Spoken-Language Access to Multimedia (SLAM): A Multimodal Interface to the World-Wide Web," Masters Thesis, Oregon Graduate Institute, Department of Computer Science & Engineering, 59 pages, Apr. 1995.  
 Julia, L. et al., "http://www.speech.sri.com/demos/ atis.html," believed to be published in: Proceedings AAAI'97: Stanford, pp. 72-76, Jul. 1997.  
 Lau, R. et al., "Webgalaxy-Integrating Spoken Language and Hypertext Navigation," believed to be published in: in Kokkinakis, G. et al., (Eds.) Eurospeech '97, Proceedings of the 5th European Conference on Speech Communication and Technology, Rhodes (Greece), Sep. 22-25, 1997: pp. 883-886, 1997.  
 Digalakis, V. et al., "Product-Code Vector Quantization of Cepstral Parameters for Speech Recognition over the WWW," believed to be published in: Proc. ICSLP '98, 4 pages, 1998.  
 Melin, H., "On Word Boundary Detection in Digit-Based Speaker Verification," believed to be published in: Workshop on Speaker Recognition and Its Commercial and Forensic Applications (RLA2C), Avignon, France, Apr. 20-23, pp. 46-49, 1998.

## US 7,050,977 B1

Page 3

- Ramaswamy, G. et al., "Compression of Acoustic Features for Speech Recognition in Network Environments," believed to be published in: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 977-980, Jun. 1998.
- Lu, B. et al., "Scalability Issues in the Real Time Protocol (RTP)," Project Report for CPSC 663 (Real Time Systems), Dept. of Computer Science, Texas A & M University, 19 pages, 1999.
- Giuliani, D. et al., "Training of HMM with Filtered Speech Material for Hands-Free Recognition," believed to be published in: Proceedings of ICASSP '99, Phoenix, USA, 4 pages, 1999.
- Digalakis, V. et al., "Quantization of Cepstral Parameters for Speech Recognition over the World Wide Web," believed to be published in: IEEE Journal on Selected Areas of Communications, 22 pages, 1999.
- Tsakalidis, S. et al., "Efficient Speech Recognition Using Subvector Quantization and Discrete-Mixture HMMs," believed to be published in: Proc. ICASSP '99, 4 pages, 1999.
- Lin, B. et al., "A Distributed Architecture for Cooperative Spoken Dialogue Agents with Coherent Dialogue State and History," believed to be published in: IEEE Automatic Speech Recognition and Understanding Workshop, Keystone, Colorado, USA, 4 pages, Dec. 1999.
- Meunier, J., "RTP Payload Format for Distributed Speech Recognition," 48th IETF AVT WG—Aug. 3, 2000, 10 pages, 2000.
- Sand Cherry Networks, SoftServer product literature, 2 pages, 2001.
- Kim, H. et al., "A Bitstream-Based Front-End for Wireless Speech Recognition on IS-136 Communications System," IEEE Transactions on Speech and Audio Processing, vol. 9, No. 5, pp. 558-568, Jul. 2001.
- Agarwal, R., Towards a PURE Spoken Dialogue System for Information Access, believed to be published in *Proceedings of the ACL/EACL Workshop on Interactive Spoken Dialog Systems: Bringing Speech and NLP Together in Real Applications*, Madrid, Spain, 1997, 9 pages.
- Ammicht, Egbert et al., "Knowledge Collection for Natural Language Spoken Dialog Systems," believed to be published in *Proc. Eurospeech*, vol. 3, p. 1375-1378, Budapest, Hungary, Sep. 1999, 4 pages.
- AT&T Corp., "Network Watson 1.0 System Overview," 1998, 4 pages.
- AT&T Corp., "AT&T Watson Advanced Speech Applications Platform," 1996, 3 pages.
- AT&T Corp., "AT&T Watson Advanced Speech Applications Platform Version 2.0," 1996, 3 pages.
- AT&T Corp., "AT&T Watson Advanced Speech Application Platform Version 2.0," 1996, 8 pages.
- Gorin, Allen, "Processing of Semantic Information in Fluently Spoken Language," believed to be published in *Proc. ICSLP*, Philadelphia, PA, Oct. 1996, 4 pages.
- Gorin, Allen et al., "How May I Help You," believed to be published in *Proc. IVTTA*, Basking Ridge, NJ, Oct. 1996, 32 pages.
- Mohru, Mehryar, "String Matching With Automata," *Nordic Journal of Computing*, 1997, 15 pages.
- Prudential News, "Prudential Pilots Revolutionary New Speech-Based Telephone Customer Service System Developed by AT&T Labs—Company Business and Marketing," Dec. 6, 1999, 3 pages.
- Riccardi, Giuseppe et al., "A spoken language system for automated call routing," believed to be published in *Proc. ICASSP '97*, 1997, 4 pages.
- Sharp, Douglas, et al., "The Watson Speech Recognition Engine," accepted by ICASSP, 1997, 9 pages.
- European Patent Office search report for EP Application No. 00977144, dated Mar. 30, 2005, 5 pages.
- Burstein, A. et al., "Using Speech Recognition In A Personal Communications System," *Proceedings of the International Conference on Communications*, Chicago, Illinois, Jun. 14-18, 1992, pp. 1717-1721.
- Digalakis, V. et al., "Quantization Of Cepstral Parameters For Speech Recognition Over The World Wide Web," *IEEE Journal on Selected Areas of Communications*, 1999, pp. 82-90.
- Kuhn, T. et al., "Hybrid In-Car Speech Recognition For Mobile Multimedia Applications," *Vehicular Technology Conference*, Houston, Texas, May 1999, pp. 2009-2013.
- L. Travis, "Handbook of Speech Pathology", Appleton-Century-Crofts, Inc., 1957, pp. 91-12-4.
- L.E. Baum, T. Petrie, "Statistical inference for probabilistic functions for finite state Markov chains", *Ann. Math. Stat.*, 37: 1554-1563, 1966.
- J.L. Flanagan, "Speech Analysis Synthesis and Perception", 2<sup>nd</sup> edition, Springer-Verlag Berlin, 1972, pp. 1-53.
- L.E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes", *Inequalities* 3: 1-8, 1972.
- Cox, Richard V. et al., "Speech and Language Processing for Next-Millennium Communications Services," *Proceedings of the IEEE*, vol. 88, No. 8, Aug. 2000, pp. 1314-1337.
- Kuhn, Roland, and Renato De Mori, "The Application of Semantic Classification Trees to Natural Language Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, No. 5, May 1995, pp. 449-460.
- Unisys Corp., "Natural Language Speech Assistant (NLSA) Capabilities Overview," NLR 3.0, Aug. 1998, Malvern, PA, 27 pages.

\* cited by examiner

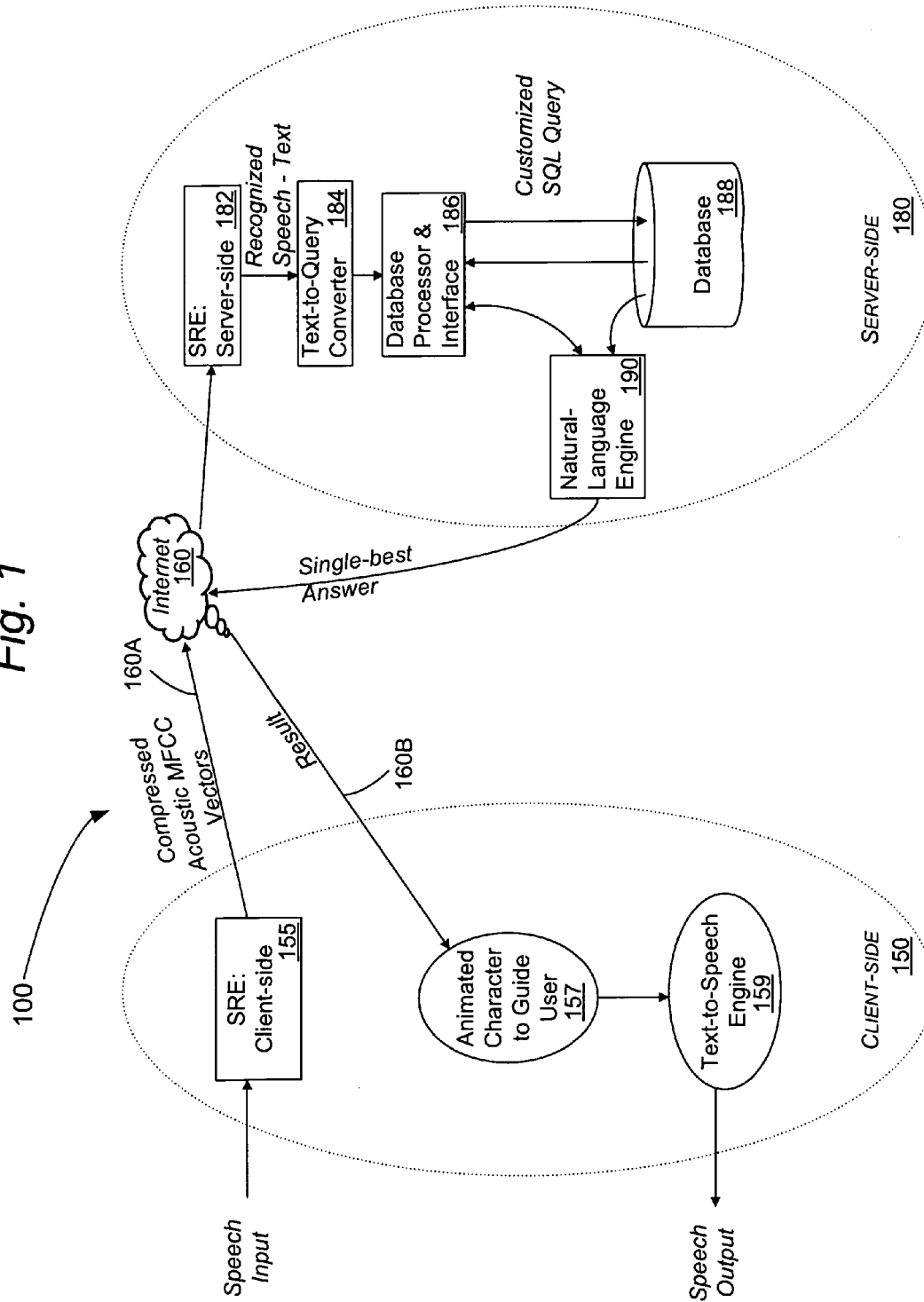
U.S. Patent

May 23, 2006

Sheet 1 of 31

US 7,050,977 B1

Fig. 1



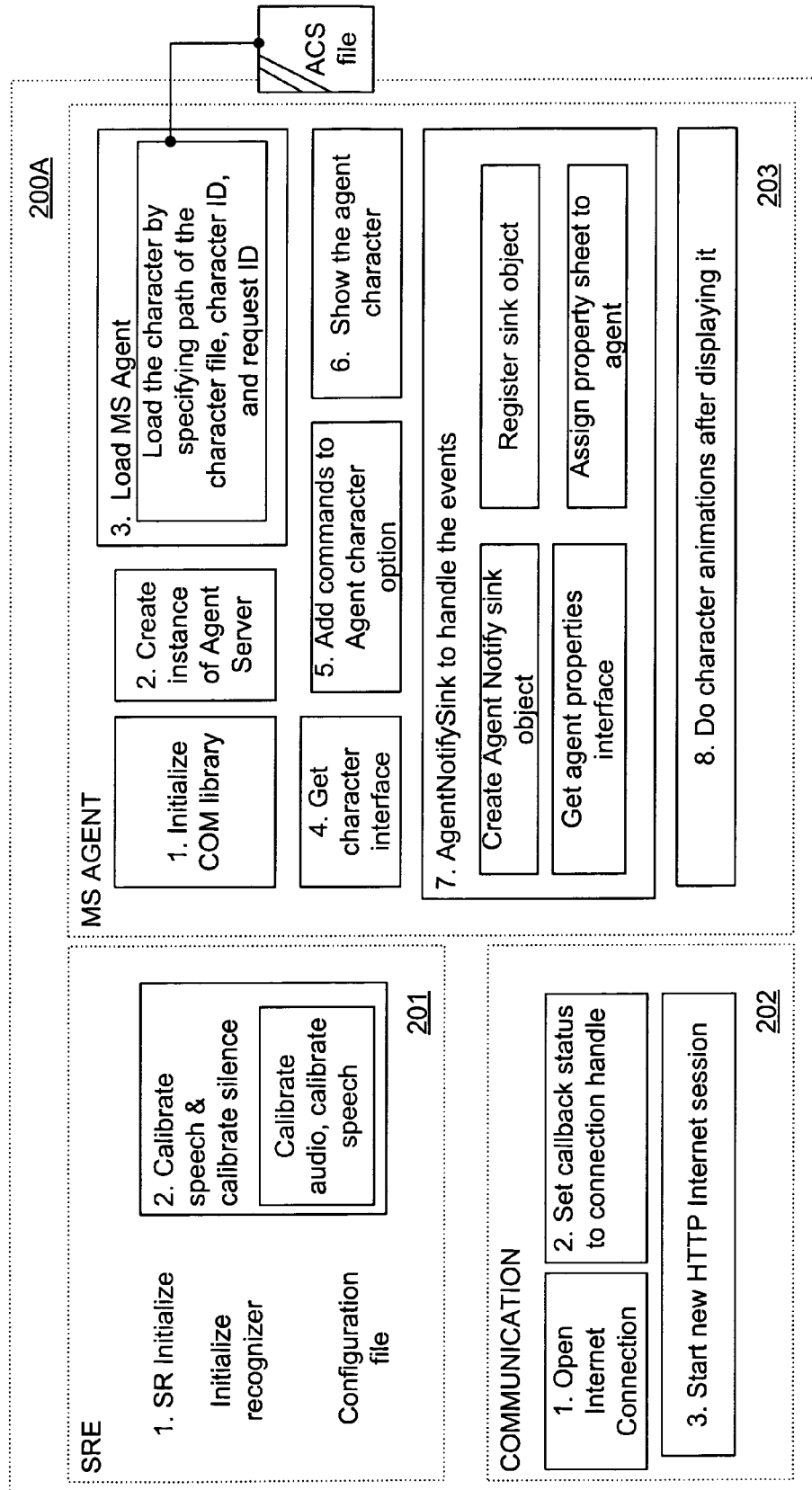
U.S. Patent

May 23, 2006

Sheet 2 of 31

US 7,050,977 B1

**Figure 2 (Page 1/3)**  
CLIENT-SIDE SYSTEM LOGIC



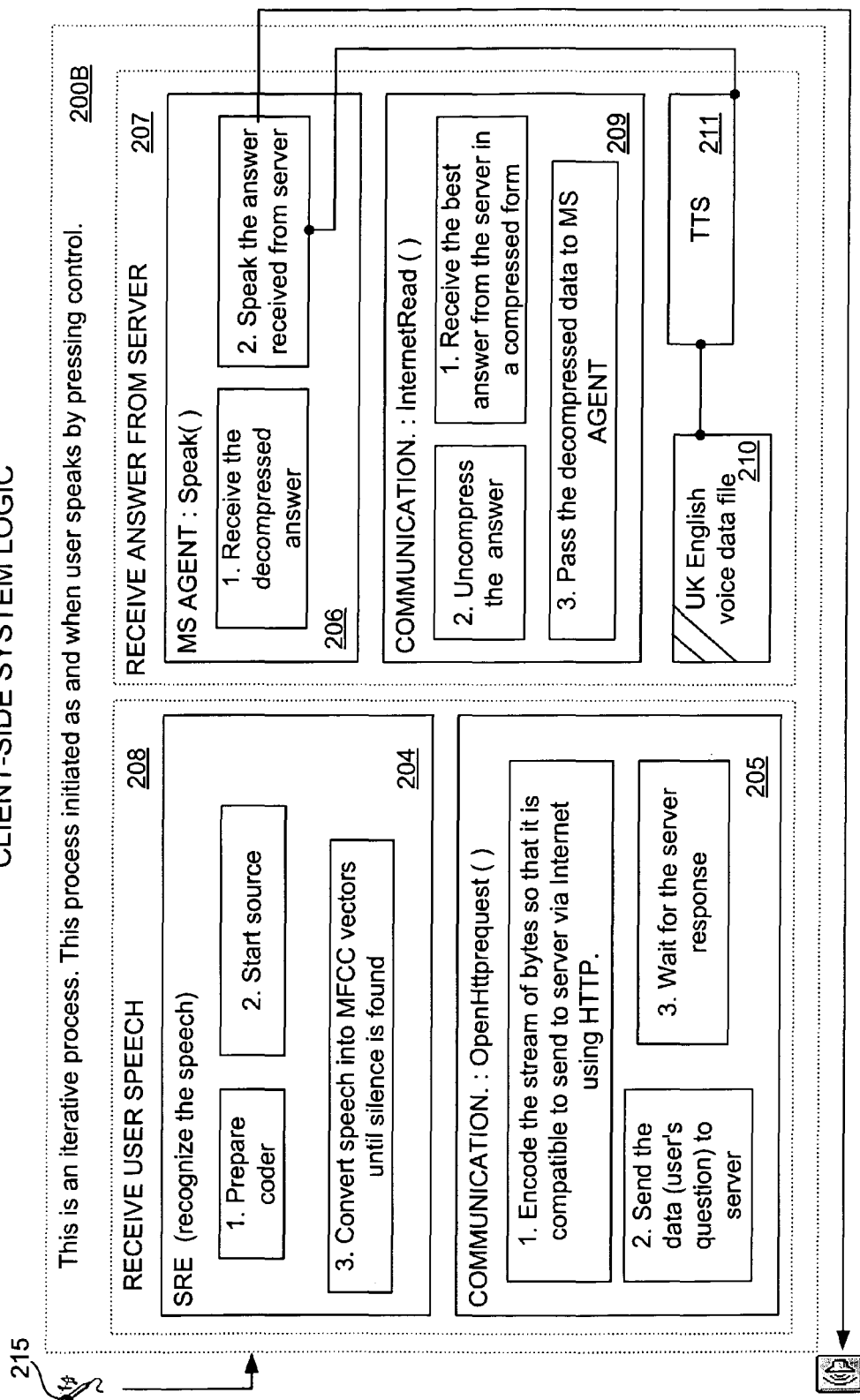
U.S. Patent

May 23, 2006

Sheet 3 of 31

US 7,050,977 B1

**Figure 2 (Page 2/3)**  
CLIENT-SIDE SYSTEM LOGIC



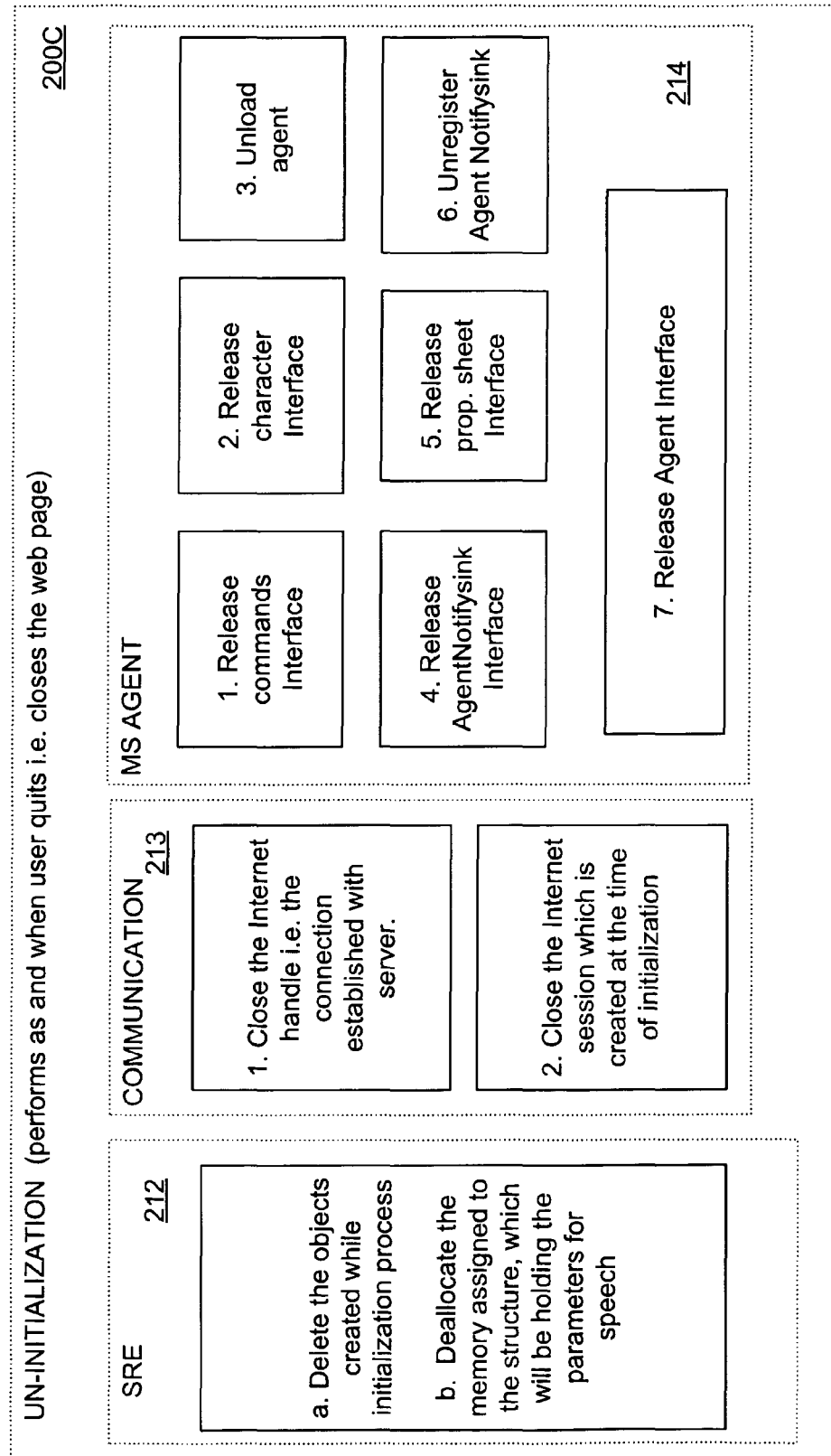
U.S. Patent

May 23, 2006

Sheet 4 of 31

US 7,050,977 B1

**Figure 2 (Page 3/3)**  
CLIENT-SIDE SYSTEM LOGIC



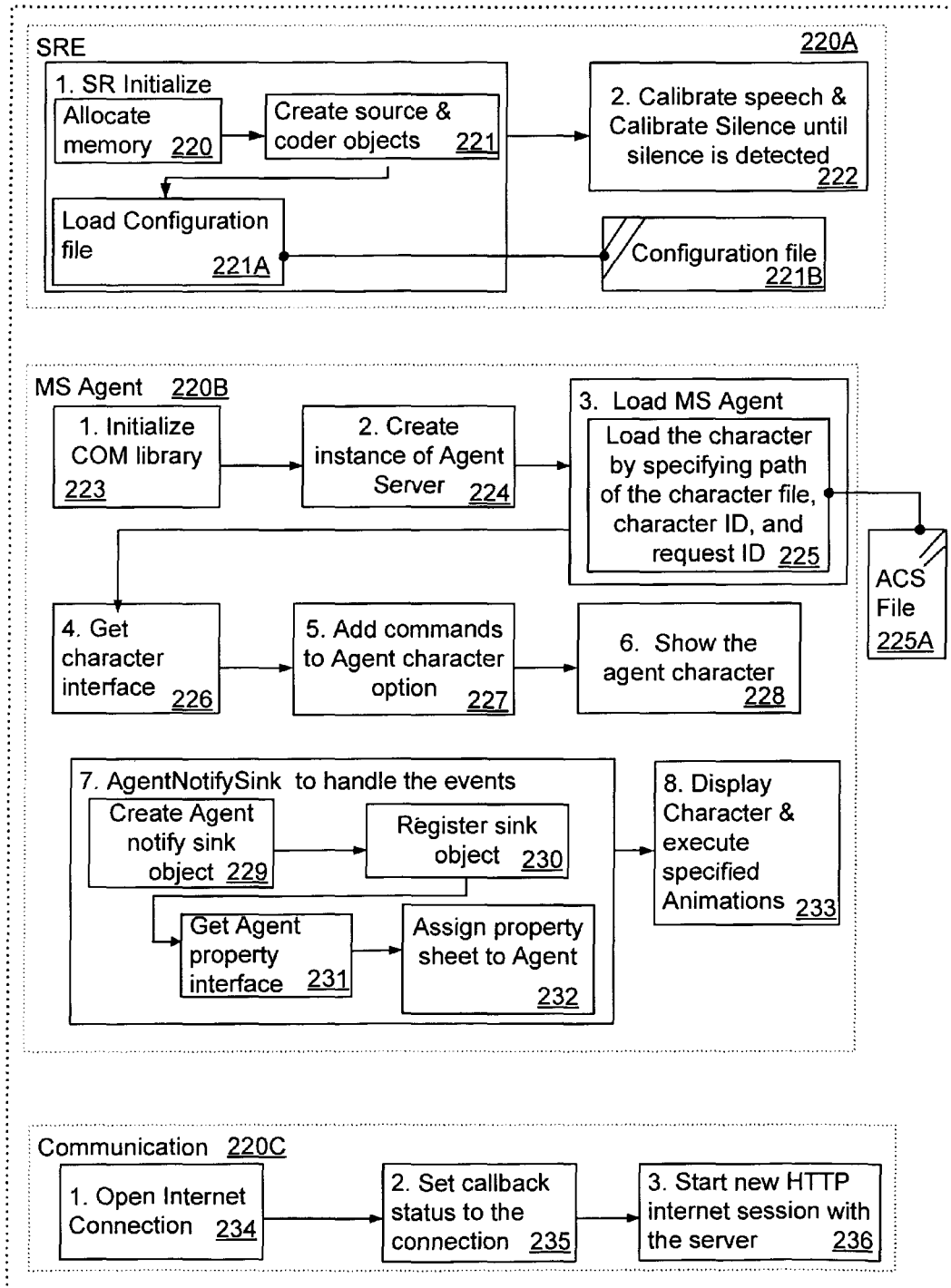


U.S. Patent

May 23, 2006

Sheet 5 of 31

US 7,050,977 B1

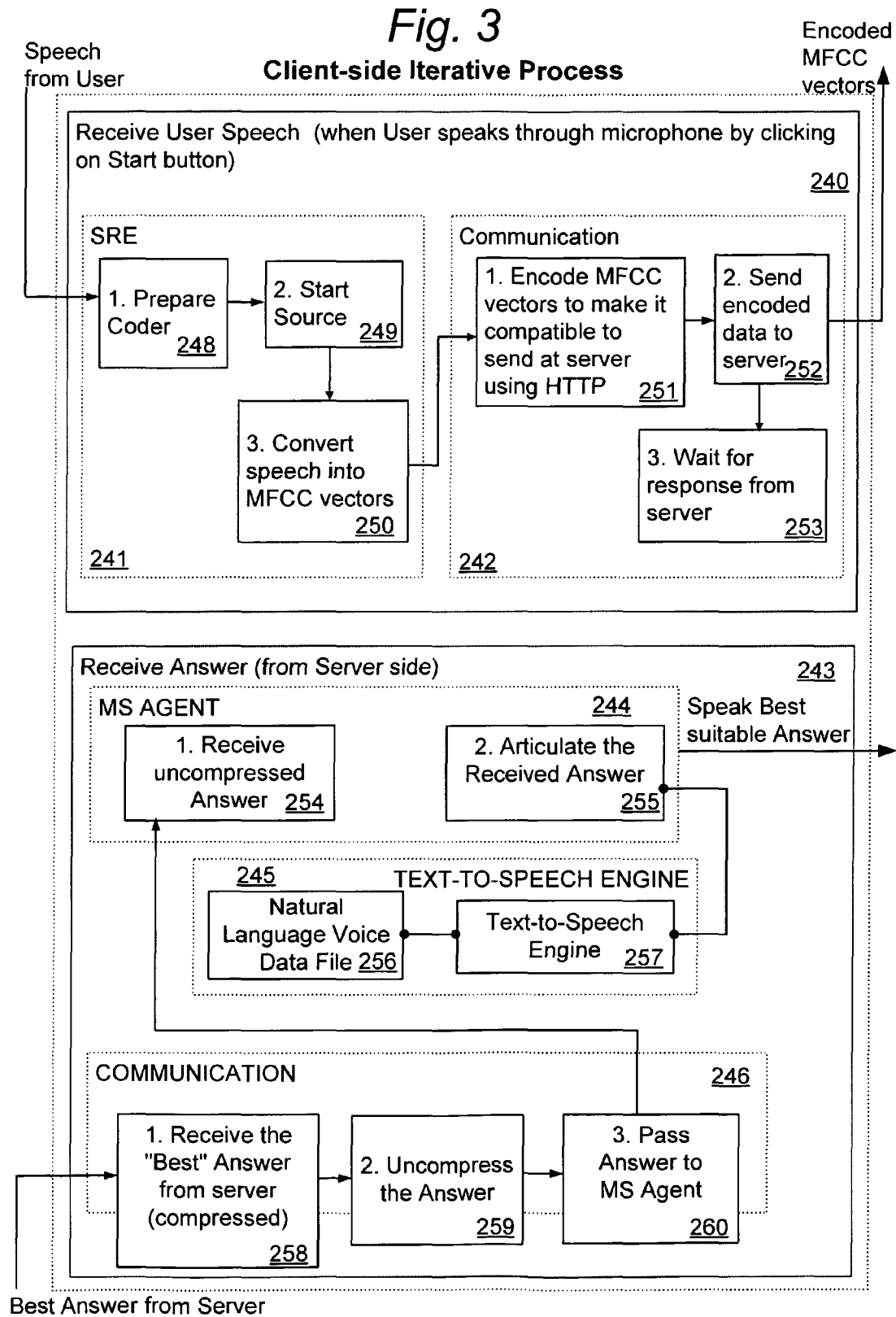
*Fig. 2-2* Client-side Initialization

U.S. Patent

May 23, 2006

Sheet 6 of 31

US 7,050,977 B1



U.S. Patent

May 23, 2006

Sheet 7 of 31

US 7,050,977 B1

*Fig. 4*  
**Client-side Un-Initialization**

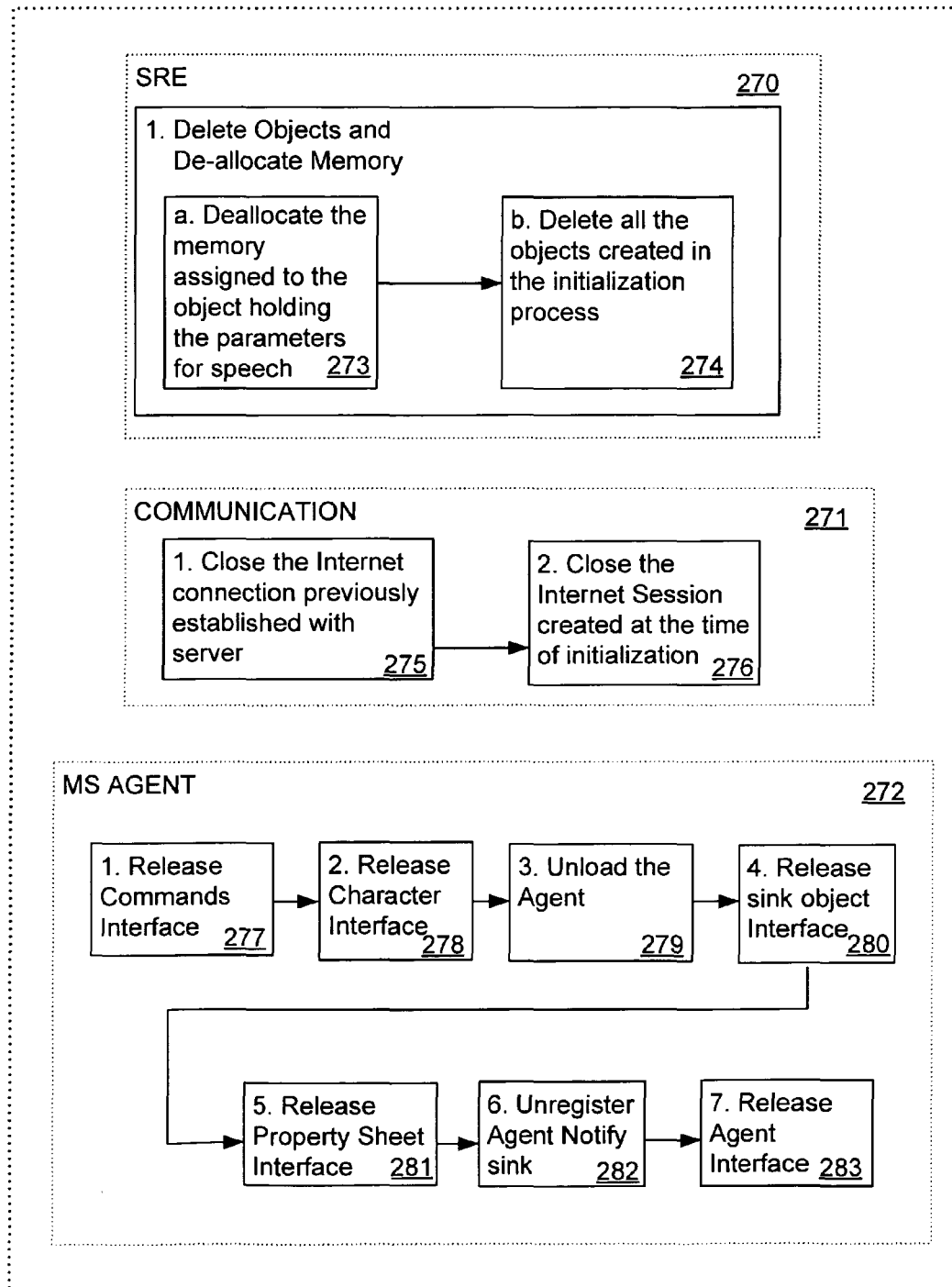
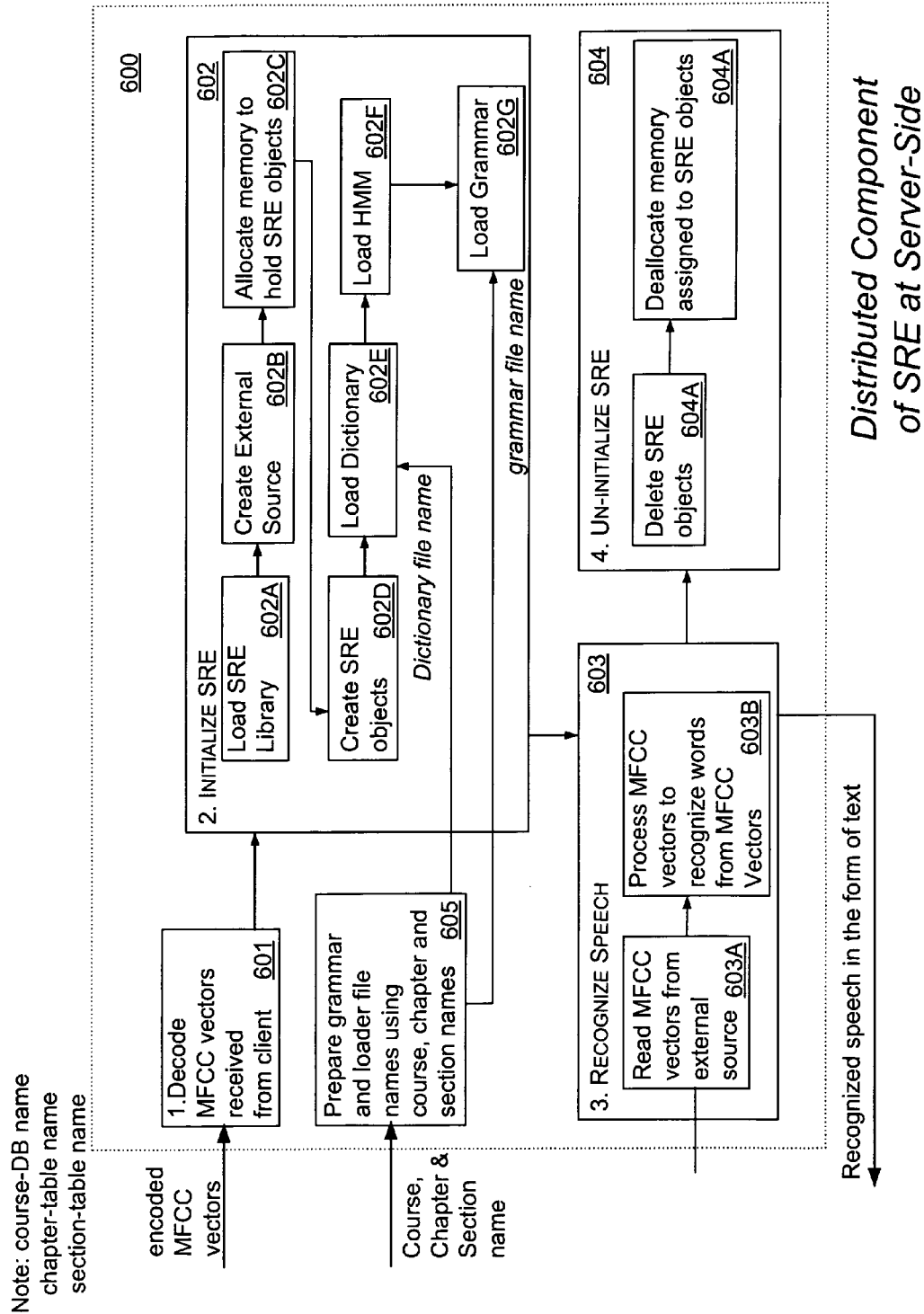


Fig. 4A



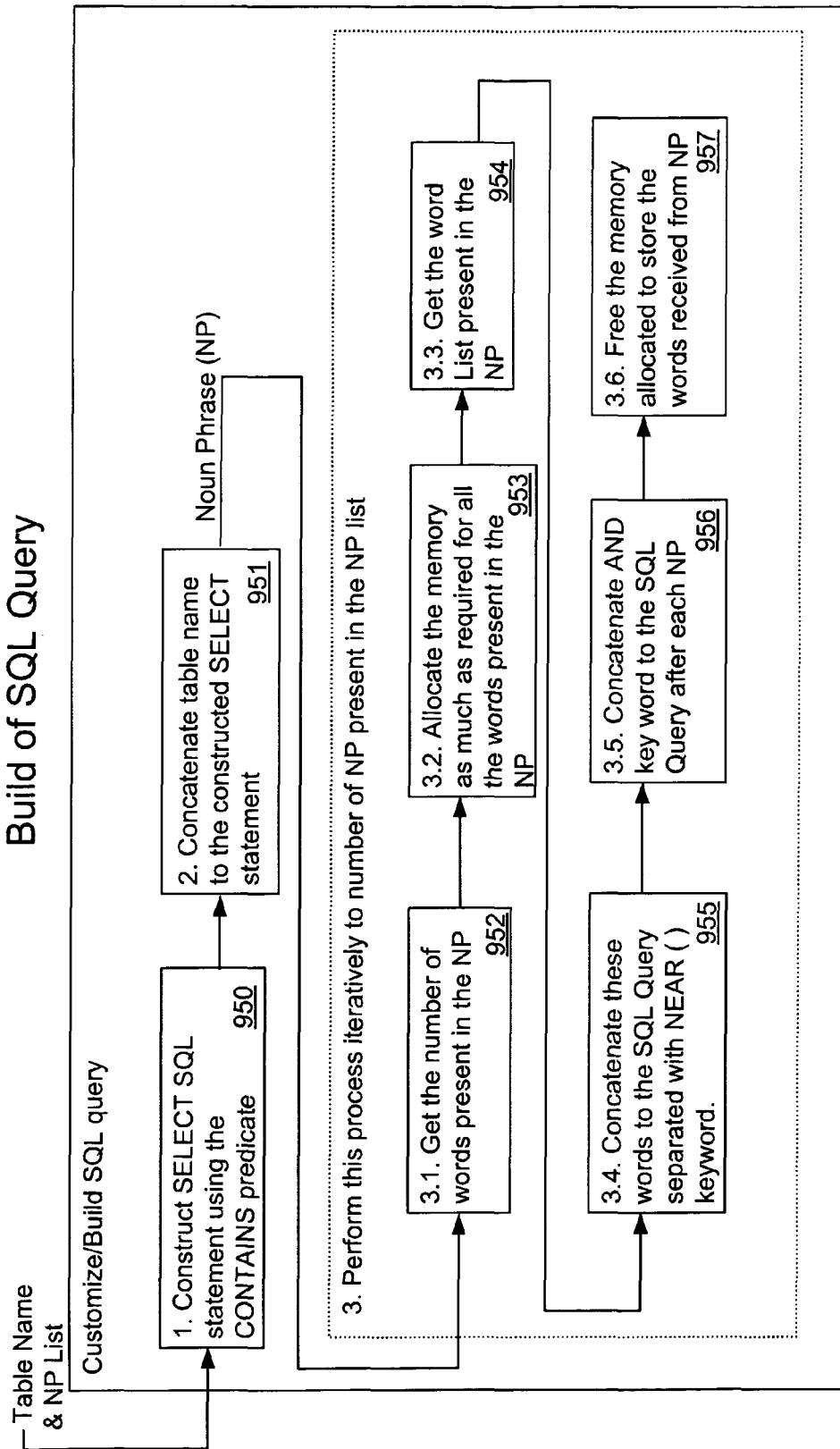
U.S. Patent

May 23, 2006

Sheet 9 of 31

US 7,050,977 B1

**Fig. 4B**  
Build of SQL Query

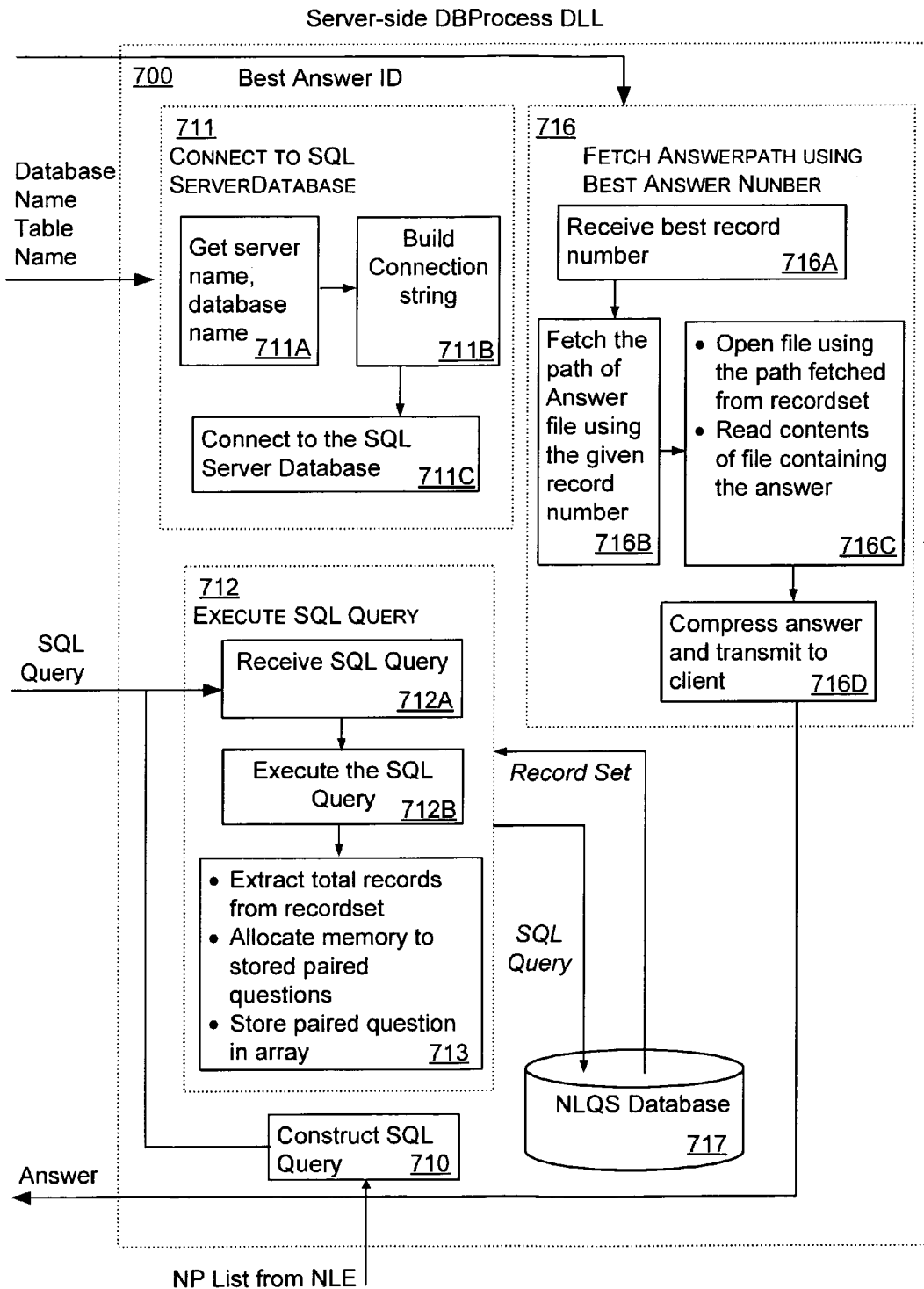


U.S. Patent

May 23, 2006

Sheet 10 of 31

US 7,050,977 B1

*Fig. 4C*

U.S. Patent

May 23, 2006

Sheet 11 of 31

US 7,050,977 B1

*Fig. 4D*

Note: PQ - Paired Question  
 NP- Noun Phrase  
 Red Line - I / O

**Interface Logic between  
 NLE and DBProcess.DLL**

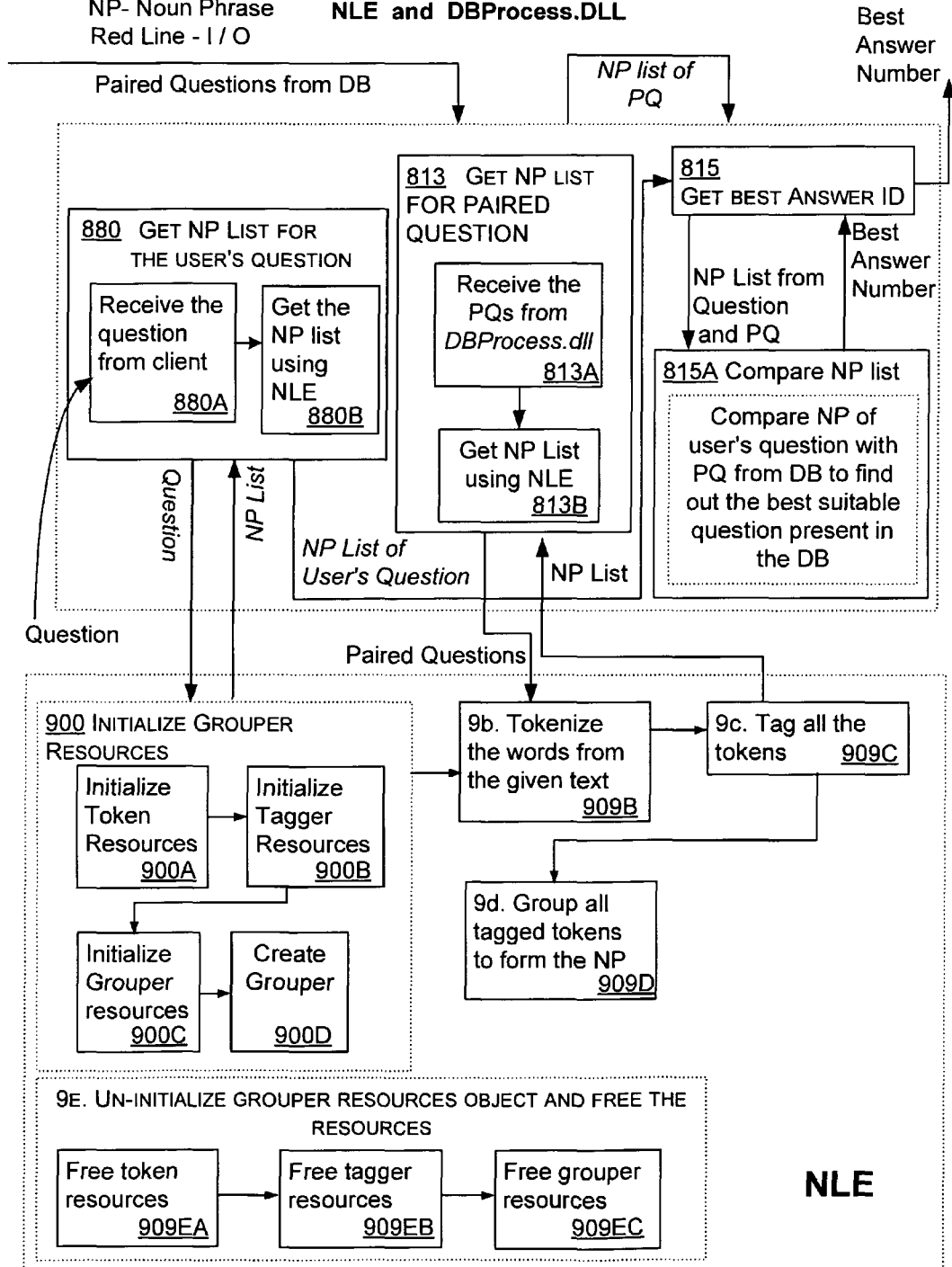


Fig. 5

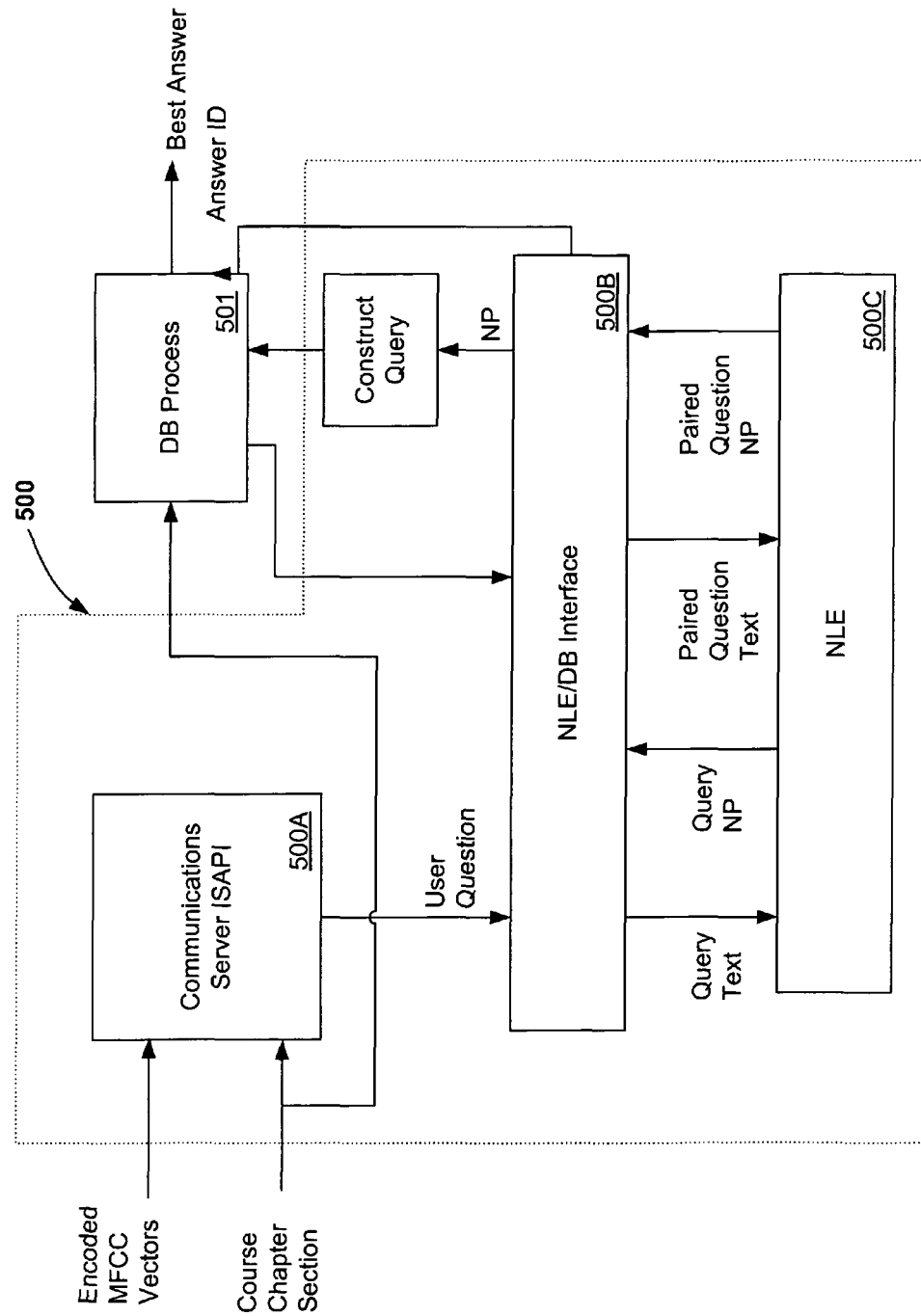




Fig. 6

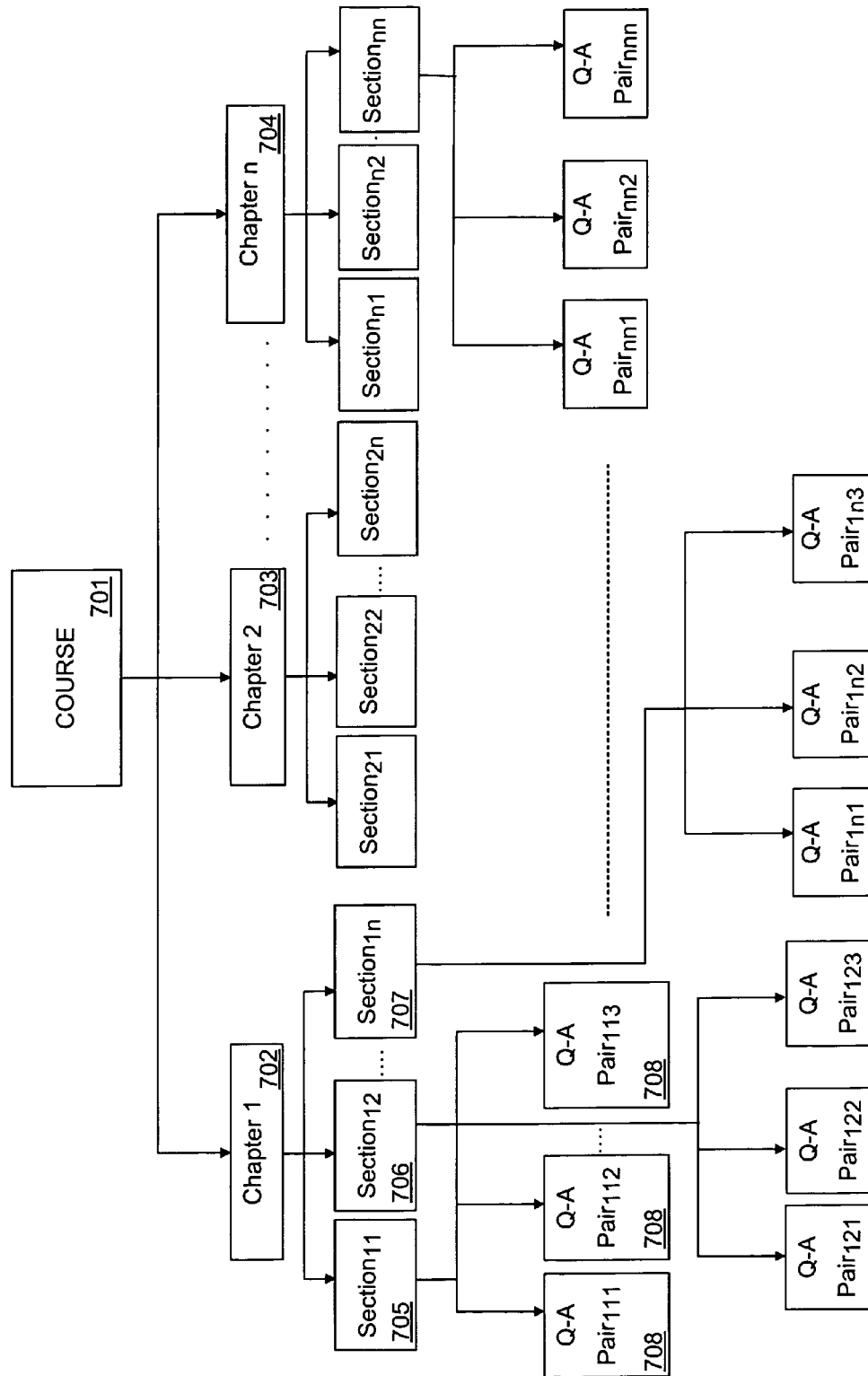


Fig. 7A

FIELD NAME <u>701A</u>	DATA TYPE <u>702A</u>	SIZE <u>703A</u>	NULL <u>704A</u>	PRIMARY KEY <u>705A</u>	INDEXED? <u>706A</u>
ChapterName <u>707A</u>	Varchar	255	No	No	Yes
SectionName <u>708A</u>	Varchar	255	No	No	Yes

U.S. Patent

May 23, 2006

Sheet 15 of 31

US 7,050,977 B1

Fig. 7B

FIELD NAME <u>720</u>	DATA TYPE <u>721</u>	SIZE <u>722</u>	NULL <u>723</u>	PRIMARY KEY <u>724</u>	INDEXED? <u>725</u>
Chapter_ID <u>726</u>	Integer		No	Yes	Yes
Answer_ID <u>727</u>	Char	5	No	UNIQUE	Yes
Section_Name <u>728</u>	Varchar	255	No	UNIQUE	Yes
Answer_Title <u>729</u>	Varchar	255	Yes	No	Yes
PairedQuestion <u>730</u>	Text	16	No	No	Yes (Full-Text)
AnswerPath <u>731</u>	Varchar	255	No	No	Yes
Creator <u>732</u>	Varchar	50	No	No	Yes
Date_of_Creation <u>733</u>	Date	-	No	No	Yes
Date_of_Modification <u>734</u>	Date	-	No	No	Yes

U.S. Patent

May 23, 2006

Sheet 16 of 31

US 7,050,977 B1

Fig. 7C

Field	<u>720</u>	Description	<u>735</u>
AnswerID	<u>727</u>	An integer - automatically incremented for user convenience	
Section_Name	<u>728</u>	Name of section to which the particular record belongs. This field along with AnswerID has to be made primary key	
Answer_Title	<u>729</u>	A short description of the answer	
PairedQuestion	<u>730</u>	Contains one or more combinations of questions for the related answer whose path is stored in the next column AnswerPath	
AnswerPath	<u>731</u>	Contains the path of text file, which contains the answer to the related questions stored in the previous column	
Creator	<u>732</u>	Name of content creator	
Date_of_Creation	<u>733</u>	Date on which content has been added	
Date_of_Modification	<u>734</u>	Date on which content has been changed or modified	

U.S. Patent

May 23, 2006

Sheet 17 of 31

US 7,050,977 B1

*Fig. 7D*

FIELD <u>740</u>	DATA TYPE <u>741</u>	SIZE <u>742</u>	NULL <u>743</u>	PRIMARY KEY <u>744</u>	INDEXED <u>745</u>
Answer_ID <u>746</u>	Char	5	No	Yes	Yes
Answer_Title <u>747</u>	Varchar	255	Yes	No	No
PairedQuestion <u>748</u>	Text	16	No	No	Yes (Full-Text)
Answer_Path <u>749</u>	Varchar	255	No	No	No
Creator <u>750</u>	Varchar	50	No	No	No
Date_of_Creation <u>751</u>	Date	-	No	No	No
Date_of_Modification <u>752</u>	Date	-	No	No	No

U.S. Patent

May 23, 2006

Sheet 18 of 31

US 7,050,977 B1

Fig. 8

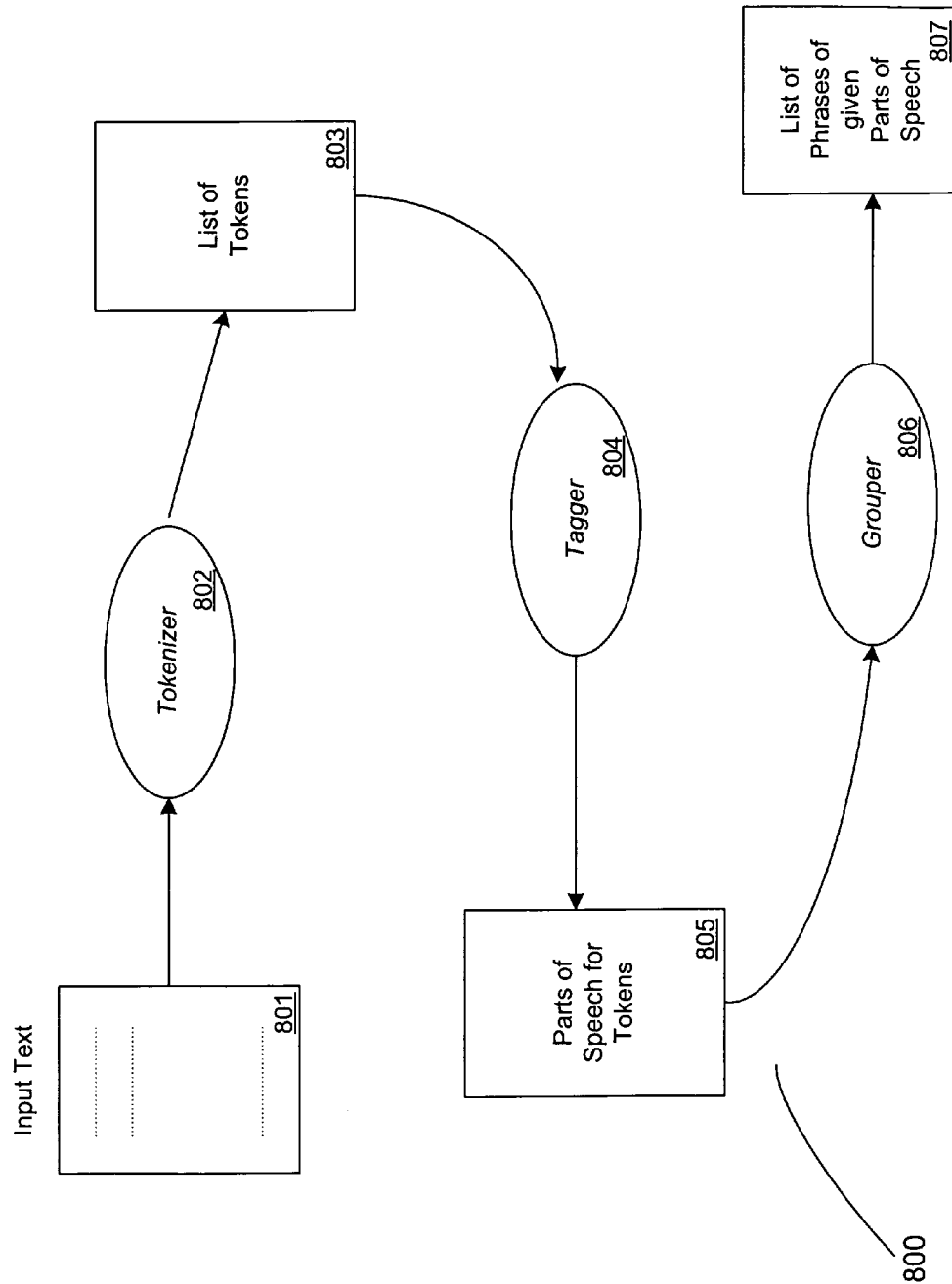


Fig. 9

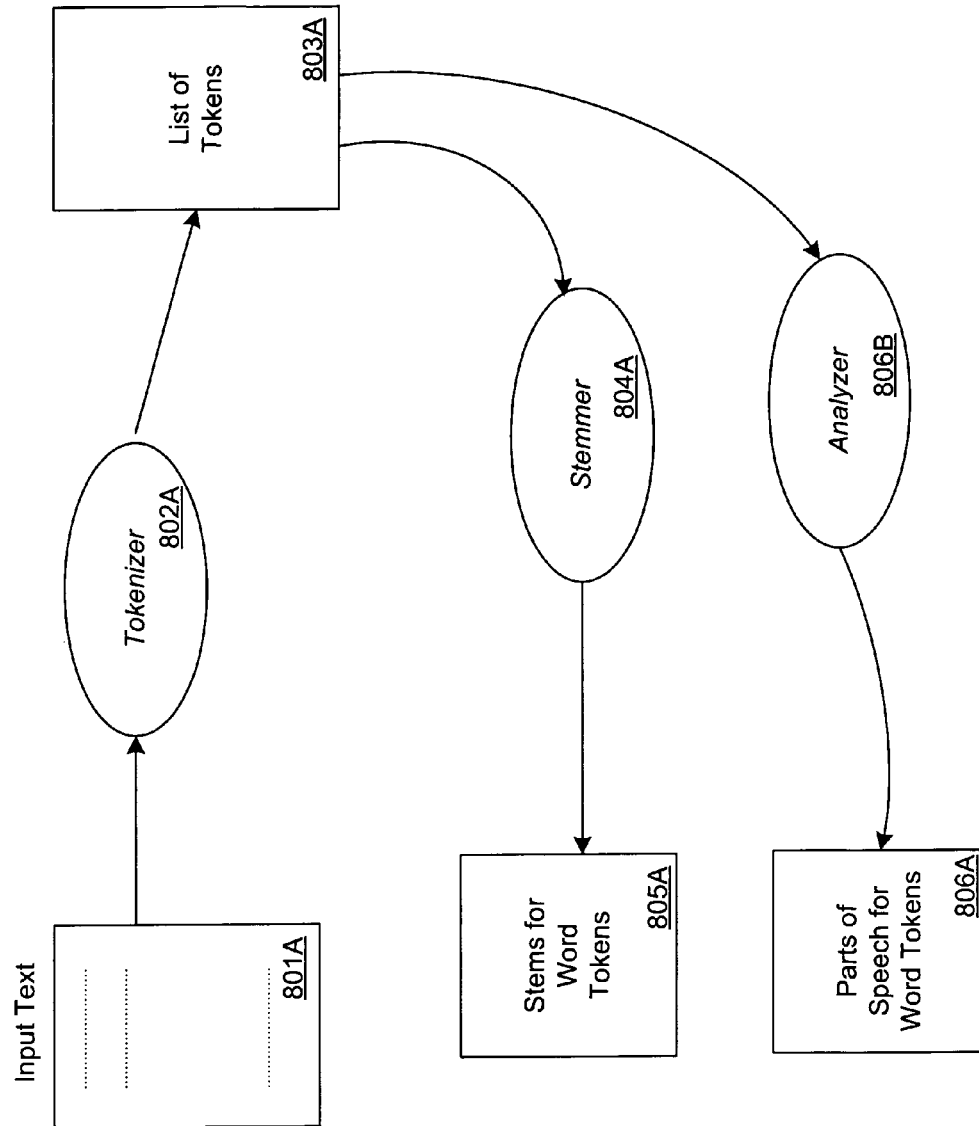
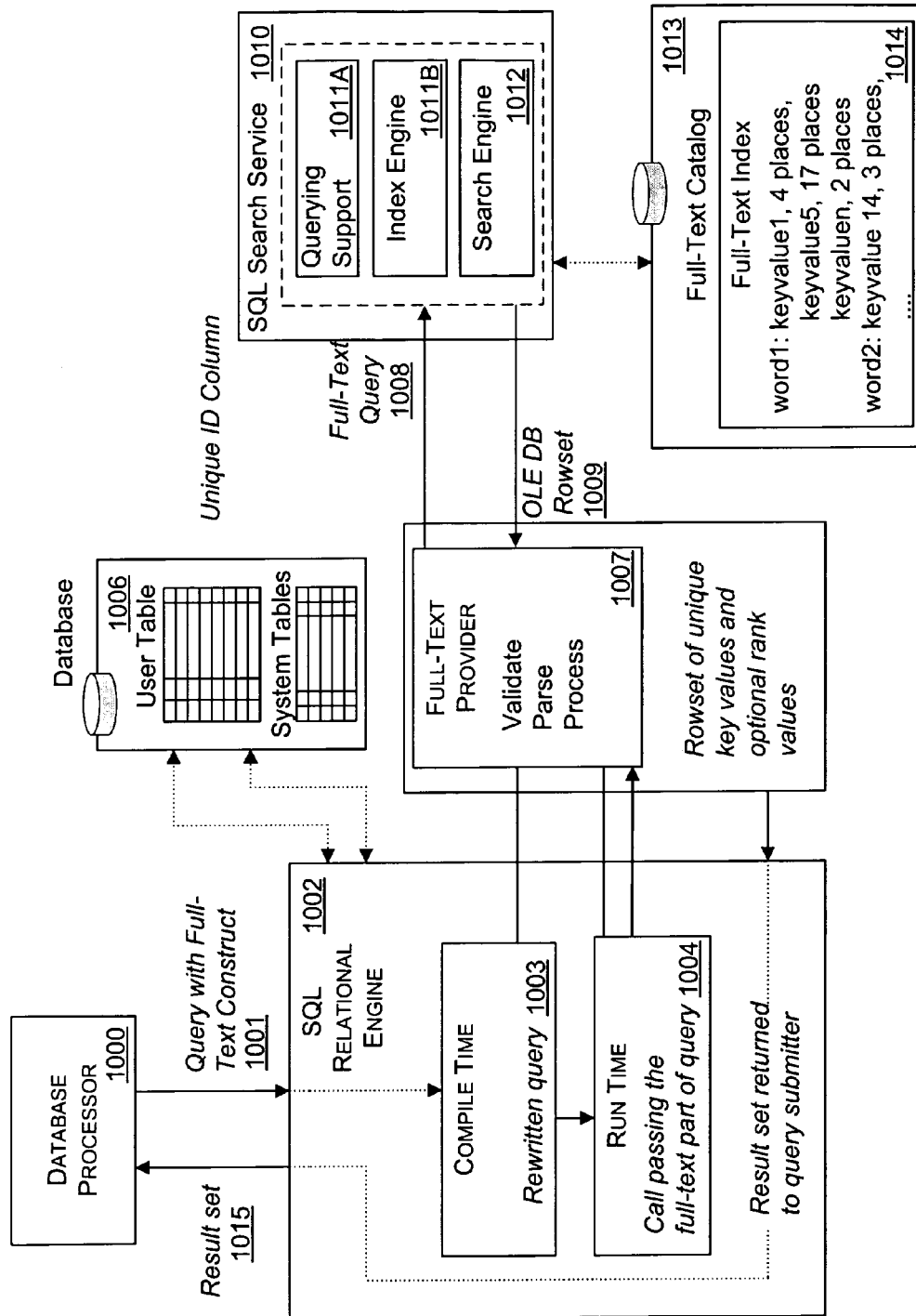


Fig. 10



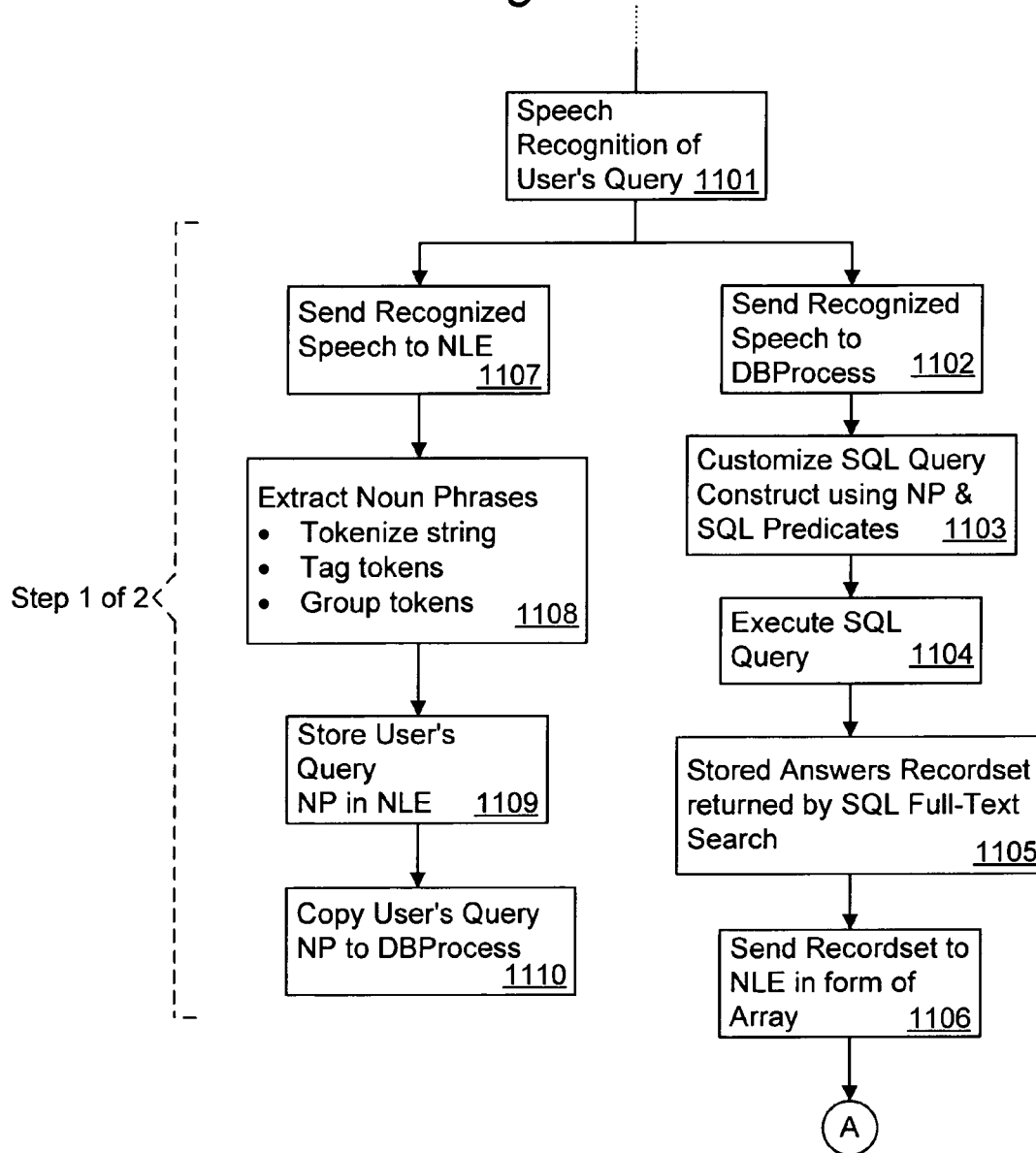


U.S. Patent

May 23, 2006

Sheet 21 of 31

US 7,050,977 B1

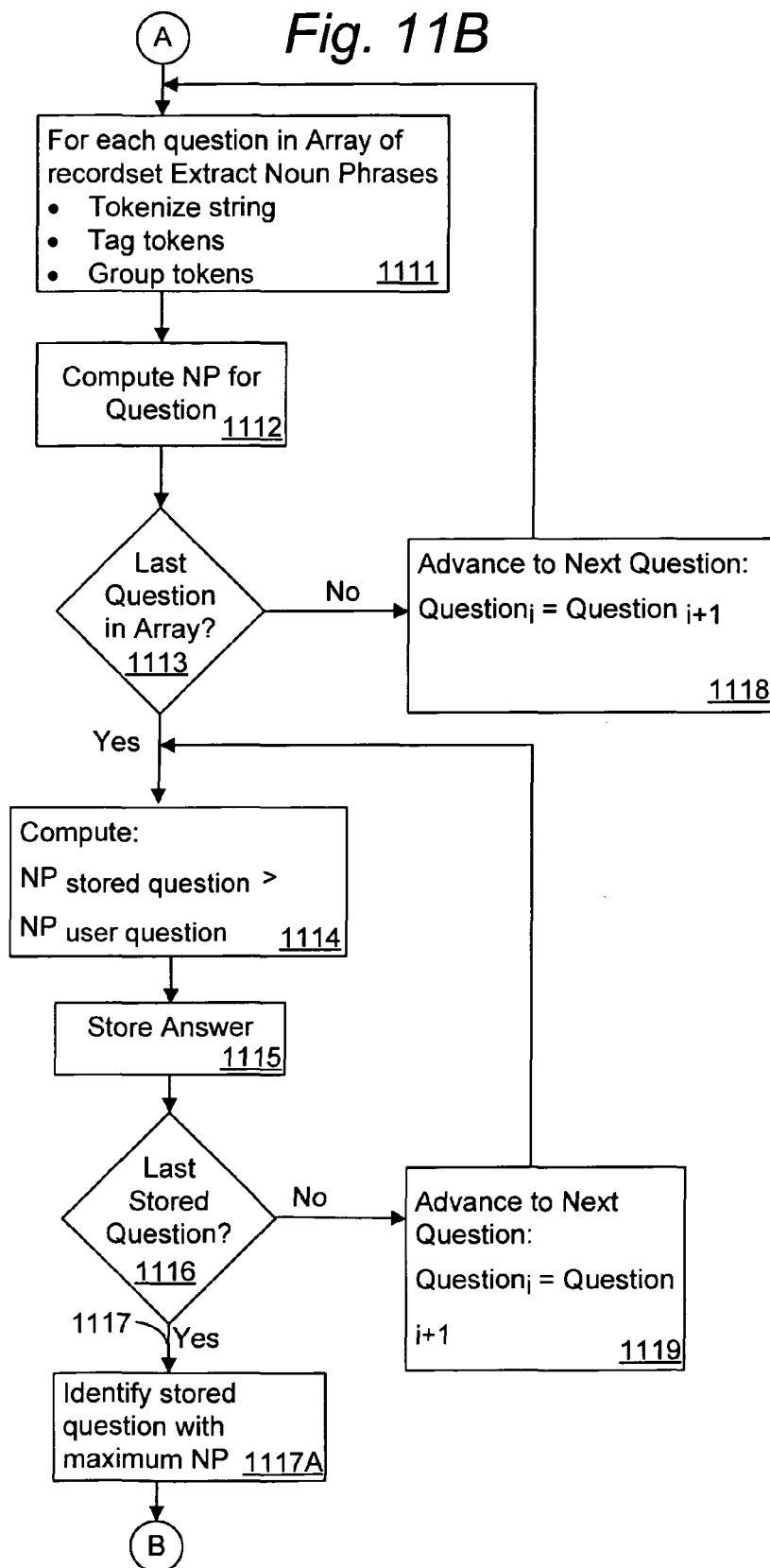
*Fig. 11A*

U.S. Patent

May 23, 2006

Sheet 22 of 31

US 7,050,977 B1



U.S. Patent

May 23, 2006

Sheet 23 of 31

US 7,050,977 B1

*Fig. 11C*

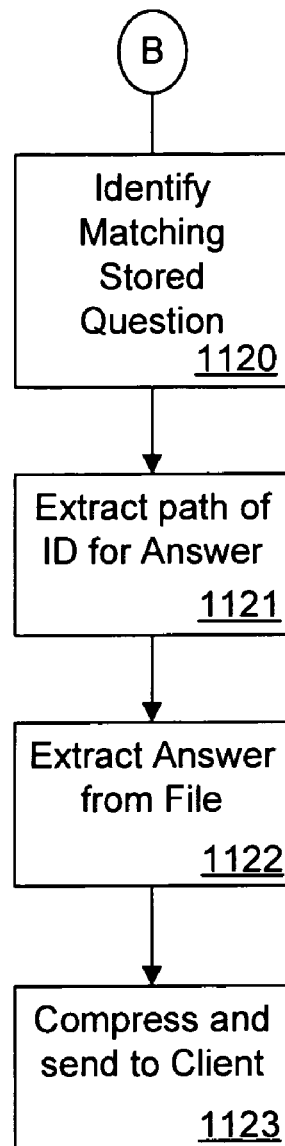
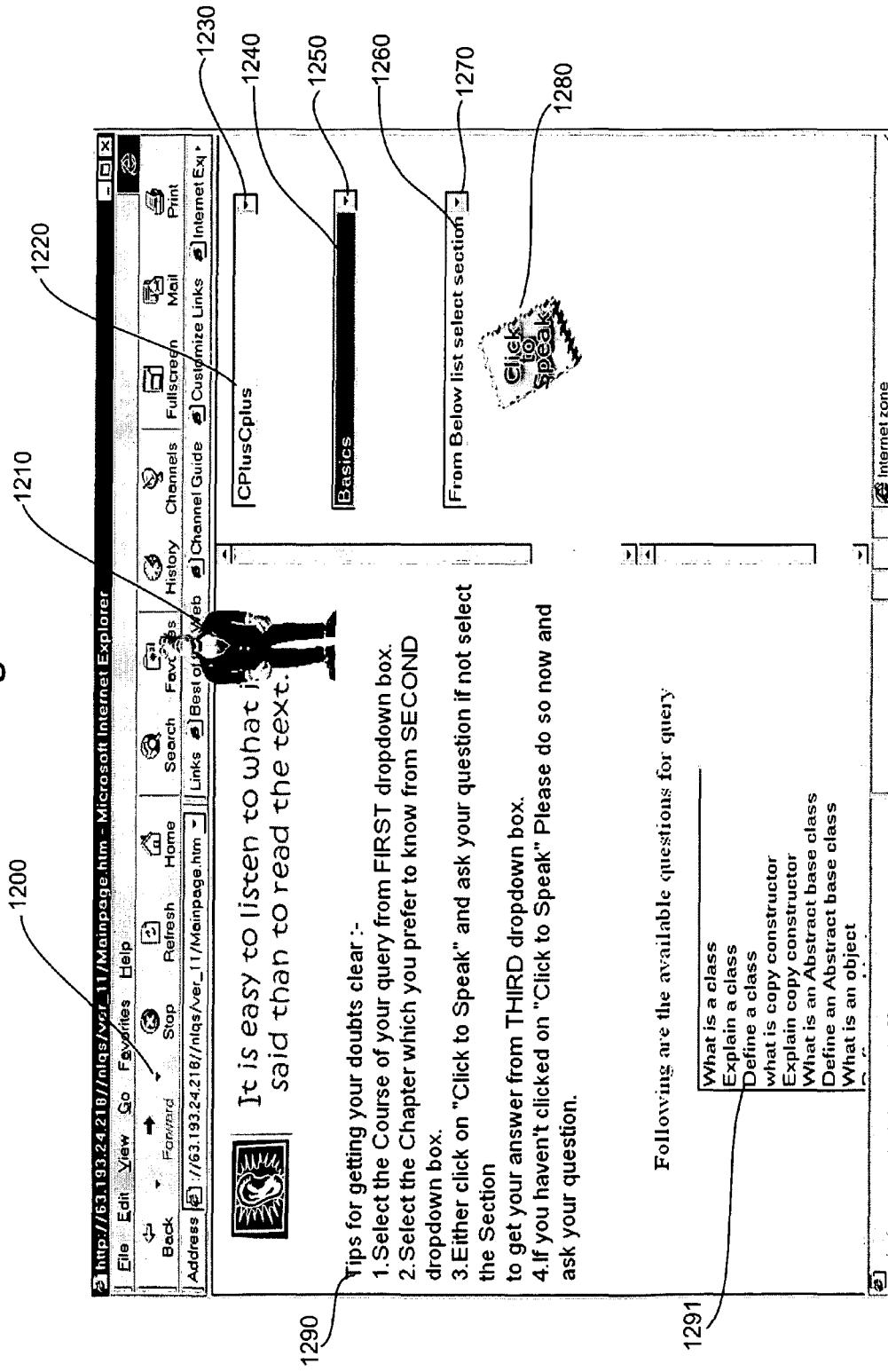


Fig. 12



U.S. Patent

May 23, 2006

Sheet 25 of 31

US 7,050,977 B1

Fig. 13

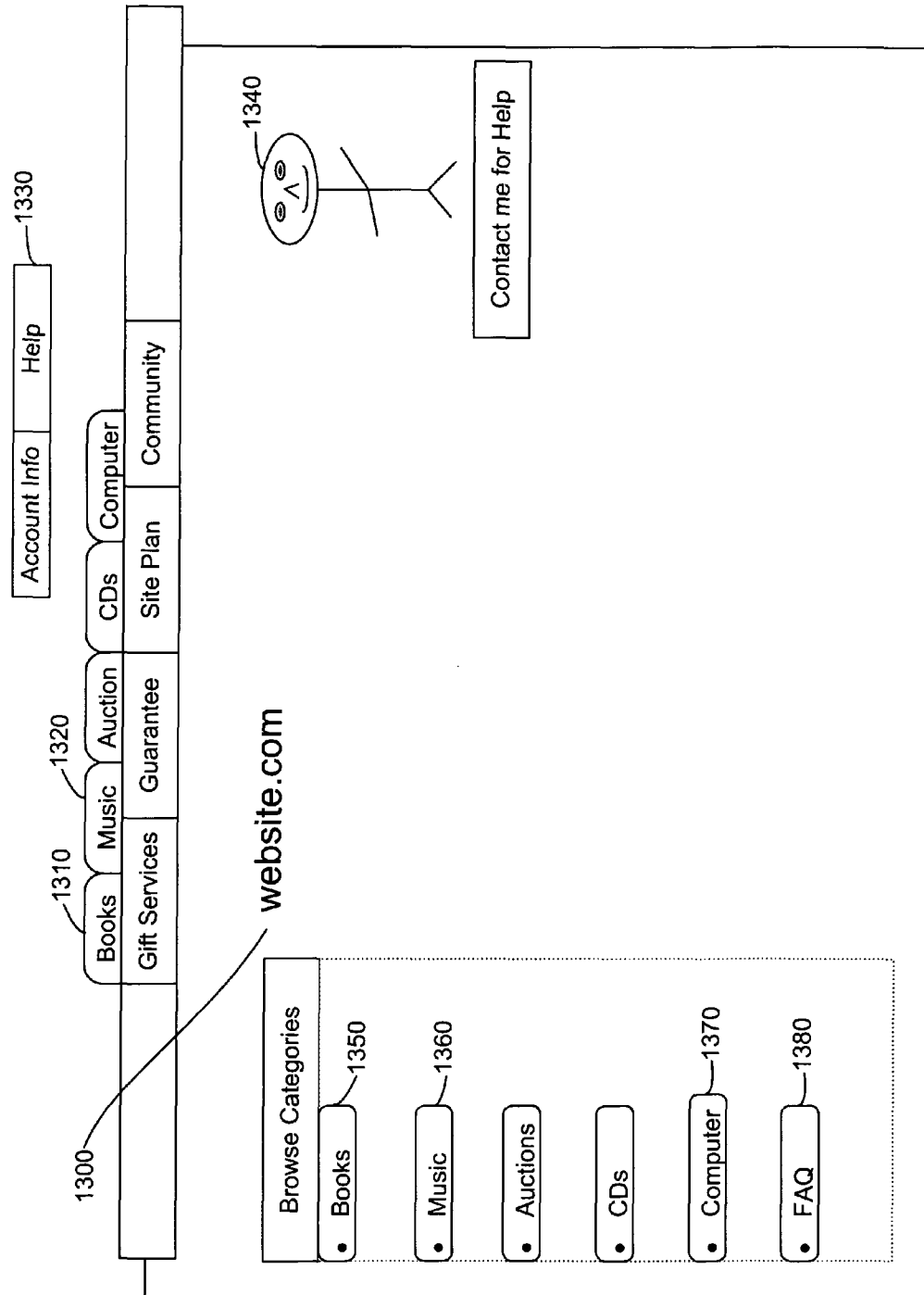
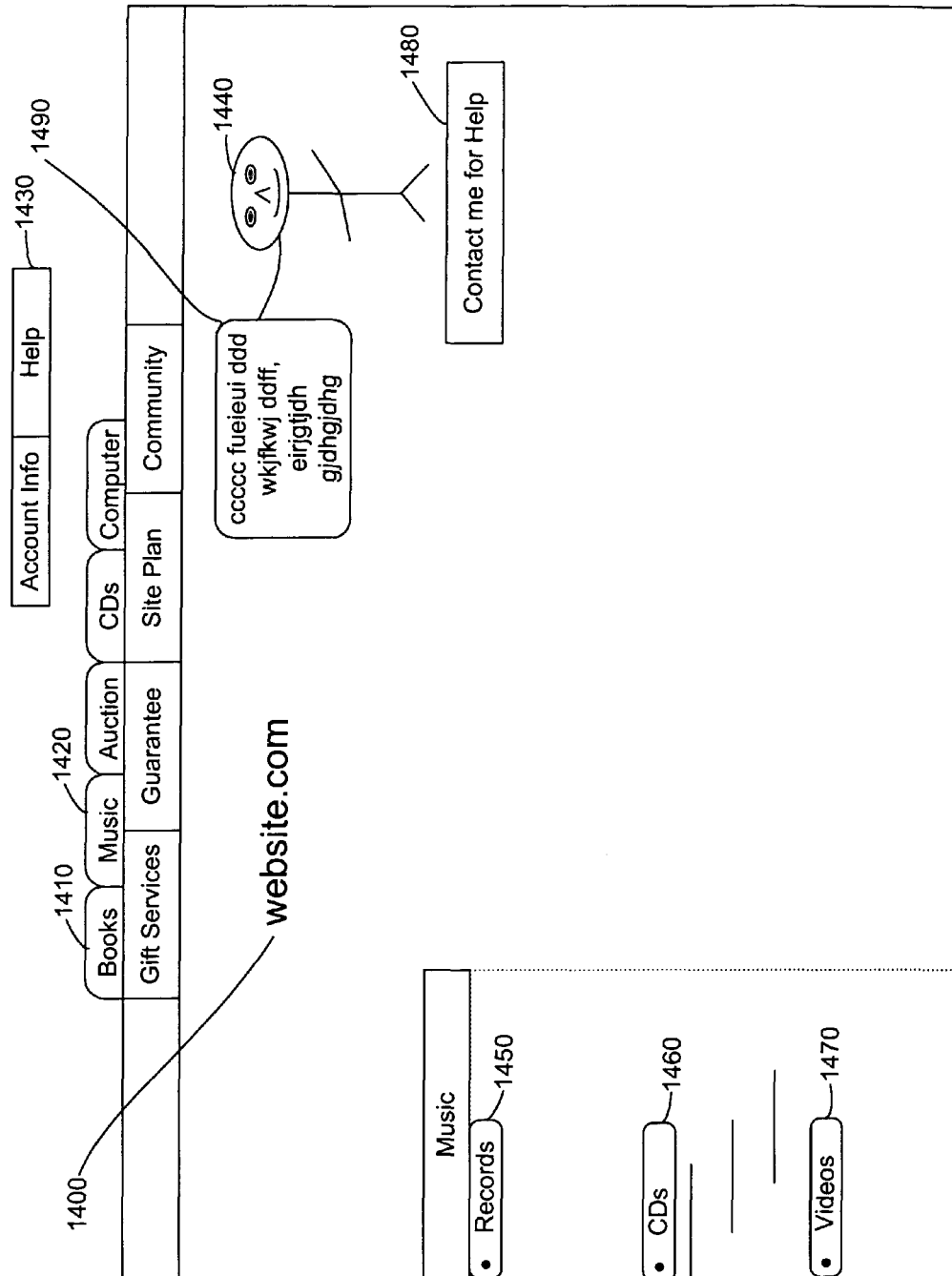


Fig. 14



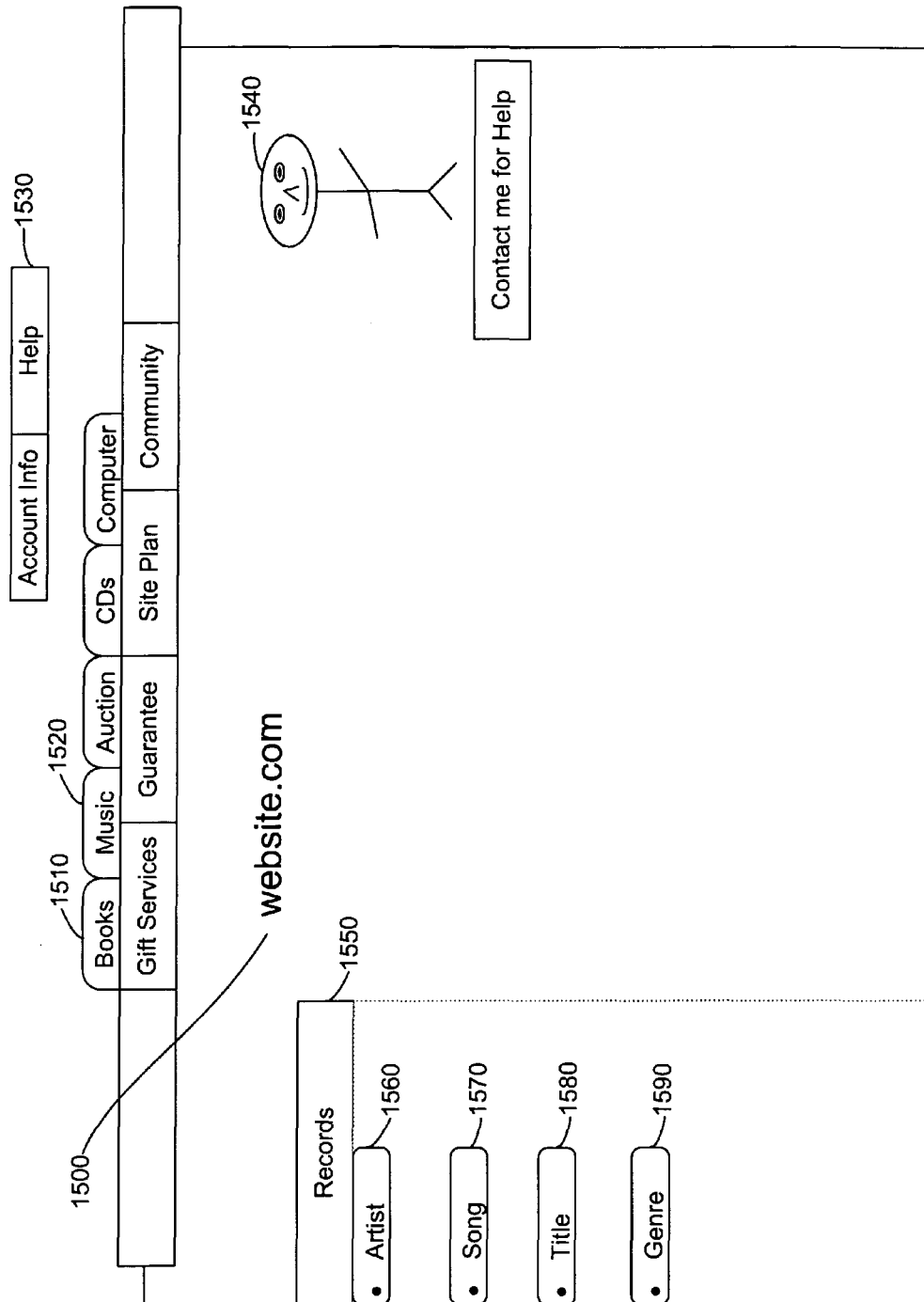
U.S. Patent

May 23, 2006

Sheet 27 of 31

US 7,050,977 B1

Fig. 15



U.S. Patent

May 23, 2006

Sheet 28 of 31

US 7,050,977 B1

Fig. 16

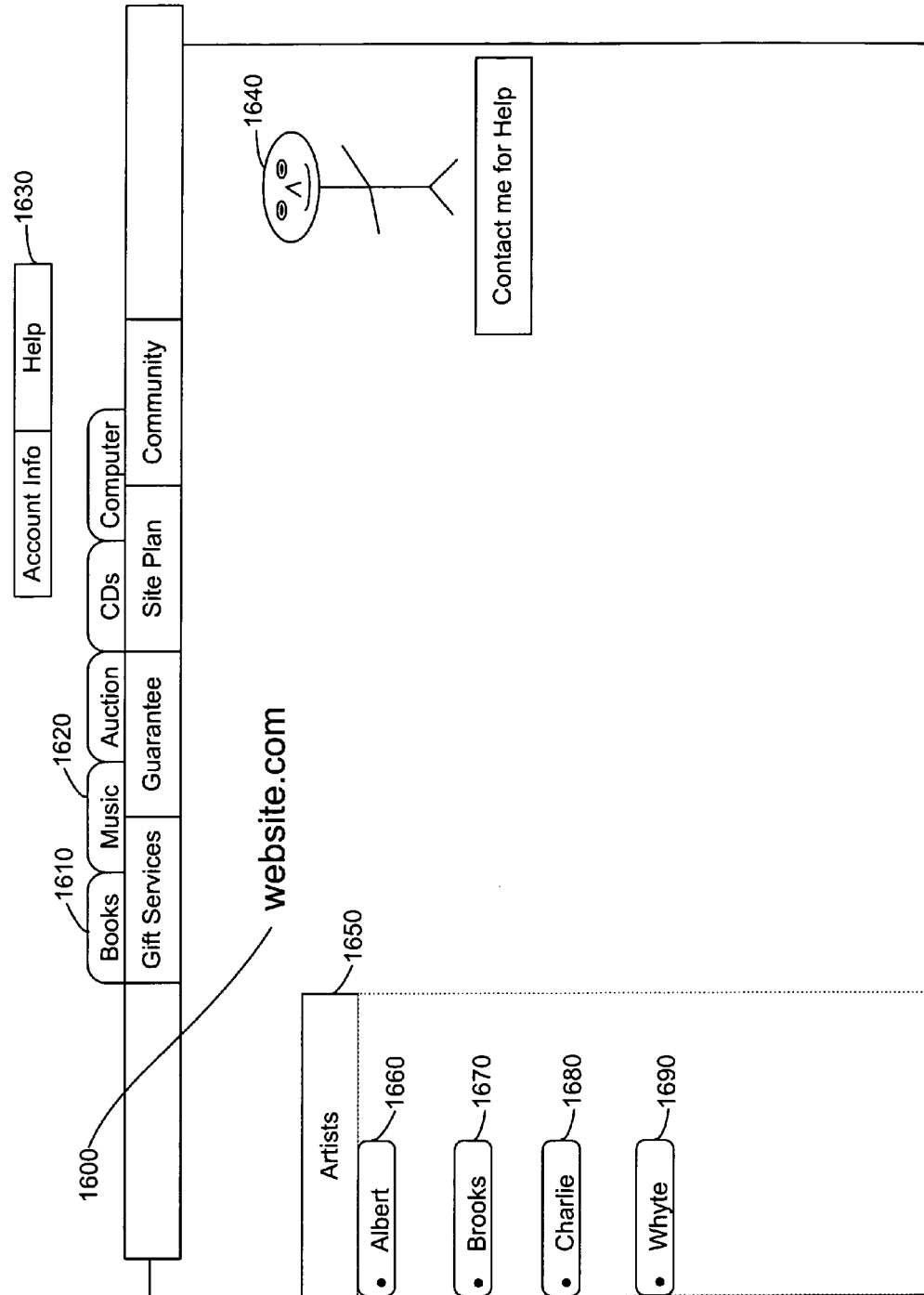
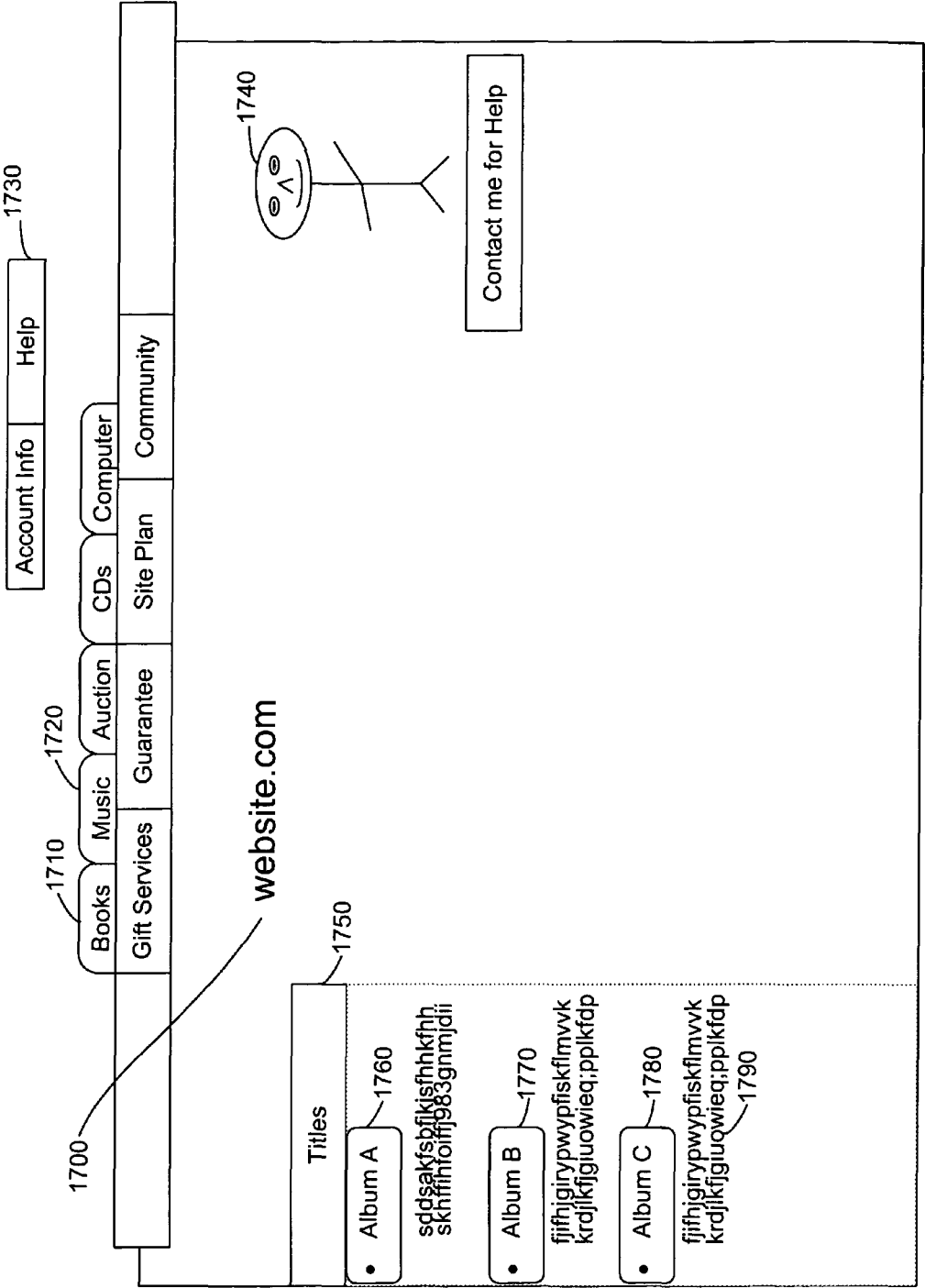




Fig. 17



U.S. Patent

May 23, 2006

Sheet 30 of 31

US 7,050,977 B1

Fig. 18 (Page 1/2)

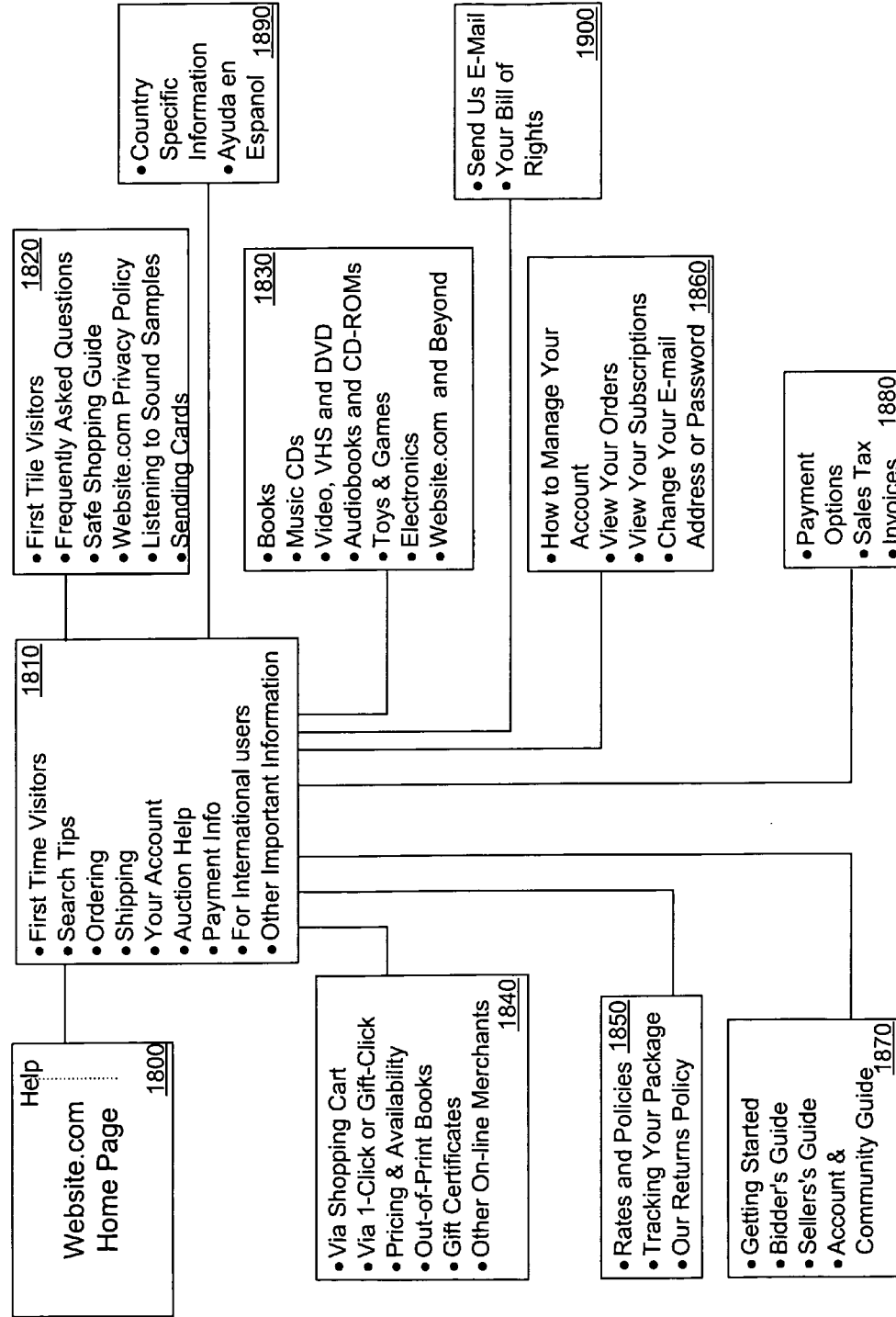
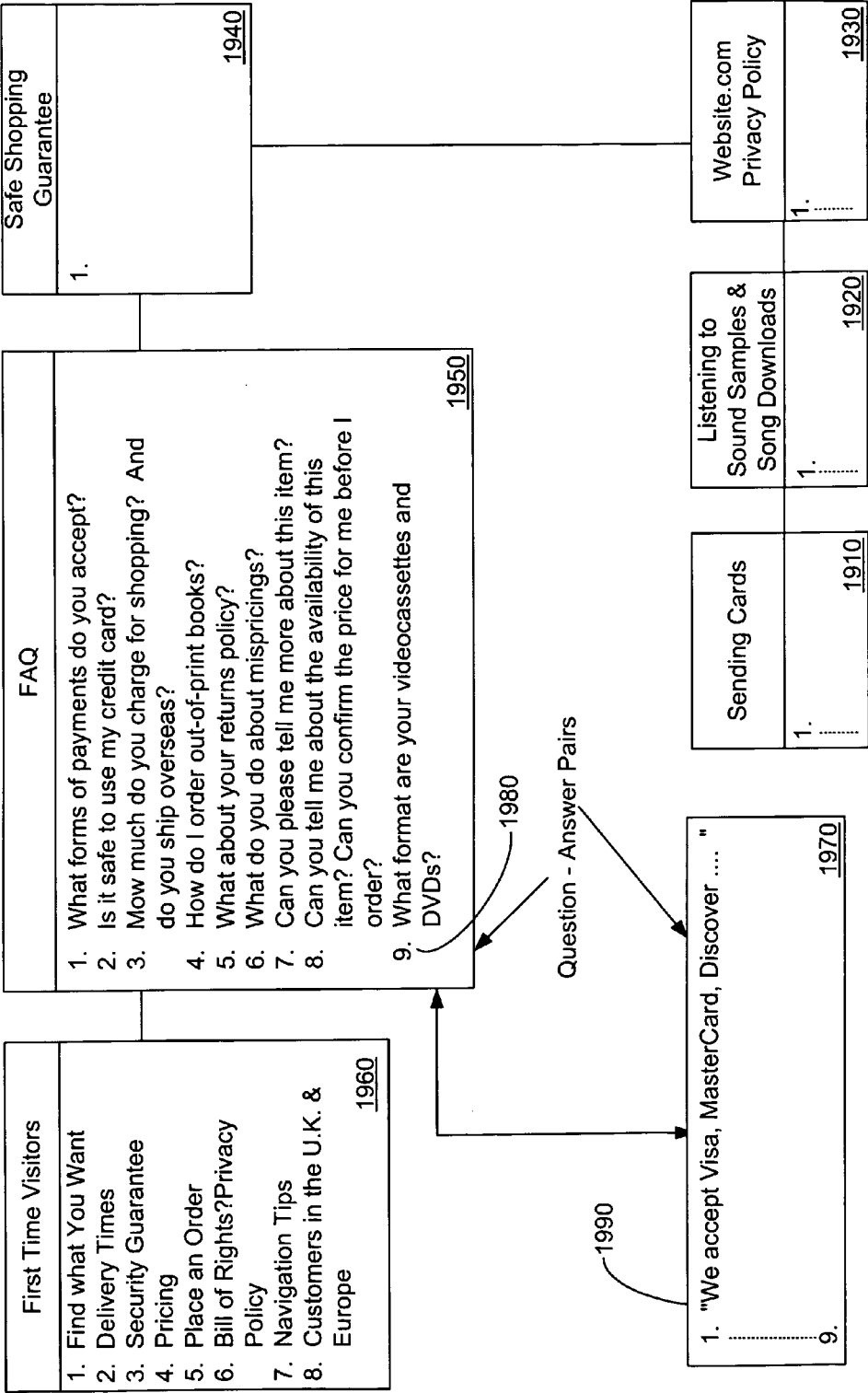


Fig. 18 (Page 2/2)



US 7,050,977 B1

1

**SPEECH-ENABLED SERVER FOR  
INTERNET WEBSITE AND METHOD**

## RELATED APPLICATIONS

The present application is related to the following applications also filed contemporaneously herewith:

- 1) Ser. No. 09/439,145 entitled Distributed Real Time Speech Recognition System;
- 2) Ser. No. 09/439,173 entitled Speech Based Learning/ Training System;
- 3) Ser. No. 09/439,060 entitled Intelligent Query Engine For Processing Voice Based Queries;

The above are incorporated by reference herein.

## FIELD OF THE INVENTION

The invention relates to a system and an interactive method for an enabling a website to have interactive, real-time speech-enabled web pages. This interactive system is especially useful when implemented for e-commerce, e-support, search engines and the like, so that a user can intelligently and easily control an internet session using a conventional browser that is enhanced to handle speech capabilities.

## BACKGROUND OF THE INVENTION

The INTERNET, and in particular, the World-Wide Web (WWW), is growing in popularity and usage for both commercial and recreational purposes, and this trend is expected to continue. This phenomenon is being driven, in part, by the increasing and widespread use of personal computer systems and the availability of low cost INTERNET access.

The emergence of inexpensive INTERNET access devices and high speed access techniques such as ADSL, cable modems, satellite modems, and the like, are expected to further accelerate the mass usage of the WWW.

Accordingly, it is expected that the number of entities offering services, products, etc., over the WWW will increase dramatically over the coming years. Until now, however, the INTERNET "experience" for users has been limited mostly to non-voice based input/output devices, such as keyboards, intelligent electronic pads, mice, trackballs, printers, monitors, etc. This presents somewhat of a bottleneck for interacting over the WWW for a variety of reasons.

First, there is the issue of familiarity. Many kinds of applications lend themselves much more naturally and fluently to a voice-based environment. For instance, most people shopping for audio recordings are very comfortable with asking a live sales clerk in a record store for information on titles by a particular author, where they can be found in the store, etc. While it is often possible to browse and search on one's own to locate items of interest, it is usually easier and more efficient to get some form of human assistance first, and, with few exceptions, this request for assistance is presented in the form of a oral query. In addition, many persons cannot or will not, because of physical or psychological barriers, use any of the aforementioned conventional I/O devices. For example, many older persons cannot easily read the text presented on WWW pages, or understand the layout/hierarchy of menus, or manipulate a mouse to make finely coordinated movements to indicate their selections. Many others are intimidated by

2

the look and complexity of computer systems, WWW pages, etc., and therefore do not attempt to use online services for this reason as well.

Thus, applications which can mimic normal human interactions are likely to be preferred by potential on-line shoppers and persons looking for information over the WWW. It is also expected that the use of voice-based systems will increase the universe of persons willing to engage in e-commerce, e-learning, etc. To date, however, there are very few systems, if any, which permit this type of interaction, and, if they do, it is very limited. For example, various commercial programs sold by IBM (VIAVOICE™) and Kurzweil (DRAGON™) permit some user control of the interface (opening, closing files) and searching (by using previously trained URLs) but they do not present a flexible solution that can be used by a number of users across multiple cultures and without time consuming voice training. Typical prior efforts to implement voice based functionality in an INTERNET context can be seen in U.S. Pat. No. 5,819,220 incorporated by reference herein.

Another issue presented by the lack of voice-based systems is efficiency. Many companies are now offering technical support over the INTERNET, and some even offer live operator assistance for such queries. While this is very advantageous (for the reasons mentioned above) it is also extremely costly and inefficient, because a real person must be employed to handle such queries. This presents a practical limit that results in long wait times for responses or high labor overheads. An example of this approach can be seen U.S. Pat. No. 5,802,526 also incorporated by reference herein. In general, a service presented over the WWW is far more desirable if it is "scalable," or, in other words, able to handle an increasing amount of user traffic with little if any perceived delay or troubles by a prospective user.

In a similar context, while remote learning has become an increasingly popular option for many students, it is practically impossible for an instructor to be able to field questions from more than one person at a time. Even then, such interaction usually takes place for only a limited period of time because of other instructor time constraints. To date, however, there is no practical way for students to continue a human-like question and answer type dialog after the learning session is over, or without the presence of the instructor to personally address such queries.

Conversely, another aspect of emulating a human-like dialog involves the use of oral feedback. In other words, many persons prefer to receive answers and information in audible form. While a form of this functionality is used by some websites to communicate information to visitors, it is not performed in a real-time, interactive question-answer dialog fashion so its effectiveness and usefulness is limited.

Yet another area that could benefit from speech-based interaction involves so-called "search" engines used by INTERNET users to locate information of interest at web sites, such as the those available at YAHOO®.com, METACRAWLER®.com, EXCITE®.com, etc. These tools permit the user to form a search query using either combinations of keywords or metacategories to search through a web page database containing text indices associated with one or more distinct web pages. After processing the user's request, therefore, the search engine returns a number of hits which correspond, generally, to URL pointers and text excerpts from the web pages that represent the closest match made by such search engine for the particular user query based on the search processing logic used by search engine. The structure and operation of such prior art search engines, including the mechanism by which they build the web page

## US 7,050,977 B1

3

database, and parse the search query, are well known in the art. To date, applicant is unaware of any such search engine that can easily and reliably search and retrieve information based on speech input from a user.

There are a number of reasons why the above environments (e-commerce, e-support, remote learning, INTERNET searching, etc.) do not utilize speech-based interfaces, despite the many benefits that would otherwise flow from such capability. First, there is obviously a requirement that the output of the speech recognizer be as accurate as possible. One of the more reliable approaches to speech recognition used at this time is based on the Hidden Markov Model (HMM)—a model used to mathematically describe any time series. A conventional usage of this technique is disclosed, for example, in U.S. Pat. No. 4,587,670 incorporated by reference herein. Because speech is considered to have an underlying sequence of one or more symbols, the HMM models corresponding to each symbol are trained on vectors from the speech waveforms. The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multi-dimensional) probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. This finite state machine changes state once every time unit, and each time  $t$  such that a state  $j$  is entered, a spectral parameter vector  $O_t$  is generated with probability density  $B_j(O_t)$ . It is only the outcome, not the state visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model. The basic theory of HMMs was published in a series of classic papers by Baum and his colleagues in the late 1960’s and early 1970’s. HMMs were first used in speech applications by Baker at Carnegie Mellon, by Jelenik and colleagues at IBM in the late 1970’s and by Steve Young and colleagues at Cambridge University, UK in the 1990’s. Some typical papers and texts are as follows:

1. L. E. Baum, T. Petrie, “Statistical inference for probabilistic functions for finite state Markov chains”, *Ann. Math. Stat.*, 37: 1554–1563, 1966
2. L. E. Baum, “An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes”, *Inequalities* 3: 1–8, 1972
3. J. H. Baker, “The dragon system—An Overview”, *IEEE Trans. on ASSP Proc.*, ASSP-23(1): 24–29, February 1975
4. F. Jeninek et al, “Continuous Speech Recognition: Statistical methods” in *Handbook of Statistics*, II, P. R. Kristnaia, Ed. Amsterdam, The Netherlands, North-Holland, 1982
5. L. R. Bahl, F. Jeninek, R. L. Mercer, “A maximum likelihood approach to continuous speech recognition”, *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-5: 179–190, 1983
6. J. D. Ferguson, “Hidden Markov Analysis: An Introduction”, in *Hidden Markov Models for Speech*, Institute of Defense Analyses, Princeton, N.J. 1980.
7. H. R. Rabiner and B. H. Juang, “Fundamentals of Speech Recognition”, Prentice Hall, 1993
8. H. R. Rabiner, “Digital Processing of Speech Signals”, Prentice Hall, 1978

More recently research has progressed in extending HMM and combining HMMs with neural networks to speech recognition applications at various laboratories. The following is a representative paper:

4

9. Nelson Morgan, Hervé Bourlard, Steve Renals, Michael Cohen and Horacio Franco (1993), Hybrid Neural Network/Hidden Markov Model Systems for Continuous Speech Recognition. *Journal of Pattern Recognition and Artificial Intelligence*, Vol. 7, No. 4 pp. 899–916. Also in I. Guyon and P. Wang editors, *Advances in Pattern Recognition Systems using Neural Networks*, Vol. 7 of a Series in Machine Perception and Artificial Intelligence. World Scientific, February 1994.

All of the above are hereby incorporated by reference. While the HMM-based speech recognition yields very good results, contemporary variations of this technique cannot guarantee a word accuracy requirement of 100% exactly and consistently, as will be required for WWW applications for all possible all user and environment conditions. Thus, although speech recognition technology has been available for several years, and has improved significantly, the technical requirements have placed severe restrictions on the specifications for the speech recognition accuracy that is required for an application that combines speech recognition and natural language processing to work satisfactorily.

In contrast to word recognition, Natural language processing (NLP) is concerned with the parsing, understanding and indexing of transcribed utterances and larger linguistic units. Because spontaneous speech contains many surface phenomena such as disfluencies,—hesitations, repairs and restarts, discourse markers such as ‘well’ and other elements which cannot be handled by the typical speech recognizer, it is the problem and the source of the large gap that separates speech recognition and natural language processing technologies. Except for silence between utterances, another problem is the absence of any marked punctuation available for segmenting the speech input into meaningful units such as utterances. For optimal NLP performance, these types of phenomena should be annotated at its input. However, most continuous speech recognition systems produce only a raw sequence of words. Examples of conventional systems using NLP are shown in U.S. Pat. Nos. 4,991,094, 5,068,789, 5,146,405 and 5,680,628, all of which are incorporated by reference herein.

Second, most of the very reliable voice recognition systems are speaker-dependent, requiring that the interface be “trained” with the user’s voice, which takes a lot of time, and is thus very undesirable from the perspective of a WWW environment, where a user may interact only a few times with a particular website. Furthermore, speaker-dependent systems usually require a large user dictionary (one for each unique user) which reduces the speed of recognition. This makes it much harder to implement a real-time dialog interface with satisfactory response capability (i.e., something that mirrors normal conversation—on the order of 3–5 seconds is probably ideal). At present, the typical shrink-wrapped speech recognition application software include offerings from IBM (VIAVOICE™) and Dragon Systems (DRAGON™). While most of these applications are adequate for dictation and other transcribing applications, they are woefully inadequate for applications such as NLQS where the word error rate must be close to 0%. In addition these offerings require long training times and are typically are non client-server configurations. Other types of trained systems are discussed in U.S. Pat. No. 5,231,670 assigned to Kurzweil, and which is also incorporated by reference herein.

Another significant problem faced in a distributed voice-based system is a lack of uniformity/control in the speech recognition process. In a typical stand-alone implementation of a speech recognition system, the entire SR engine runs on



## US 7,050,977 B1

5

a single client. A well-known system of this type is depicted in U.S. Pat. No. 4,991,217 incorporated by reference herein. These clients can take numerous forms (desktop PC, laptop PC, PDA, etc.) having varying speech signal processing and communications capability. Thus, from the server side perspective, it is not easy to assure uniform treatment of all users accessing a voice-enabled web page, since such users may have significantly disparate word recognition and error rate performances. While a prior art reference to Gould et al.—U.S. Pat. No. 5,915,236—discusses generally the notion of tailoring a recognition process to a set of available computational resources, it does not address or attempt to solve the issue of how to optimize resources in a distributed environment such as a client-server model. Again, to enable such voice-based technologies on a wide-spread scale it is far more preferable to have a system that harmonizes and accounts for discrepancies in individual systems so that even the thinnest client is supportable, and so that all users are able to interact in a satisfactory manner with the remote server running the e-commerce, e-support and/or remote learning application.

Two references that refer to a distributed approach for speech recognition include U.S. Pat. Nos. 5,956,683 and 5,960,399 incorporated by reference herein. In the first of these, U.S. Pat. No. 5,956,683—Distributed Voice Recognition System (assigned to Qualcomm) an implementation of a distributed voice recognition system between a telephony-based handset and a remote station is described. In this implementation, all of the word recognition operations seem to take place at the handset. This is done since the patent describes the benefits that result from locating of the system for acoustic feature extraction at the portable or cellular phone in order to limit degradation of the acoustic features due to quantization distortion resulting from the narrow bandwidth telephony channel. This reference therefore does not address the issue of how to ensure adequate performance for a very thin client platform. Moreover, it is difficult to determine, how, if at all, the system can perform real-time word recognition, and there is no meaningful description of how to integrate the system with a natural language processor.

The second of these references—U.S. Pat. No. 5,960,399—Client/Server Speech Processor/Recognizer (assigned to GTE) describes the implementation of a HMM-based distributed speech recognition system. This reference is not instructive in many respects, however, including how to optimize acoustic feature extraction for a variety of client platforms, such as by performing a partial word recognition process where appropriate. Most importantly, there is only a description of a primitive server-based recognizer that only recognizes the user's speech and simply returns certain keywords such as the user's name and travel destination to fill out a dedicated form on the user's machine. Also, the streaming of the acoustic parameters does not appear to be implemented in real-time as it can only take place after silence is detected. Finally, while the reference mentions the possible use of natural language processing (column 9) there is no explanation of how such function might be implemented in a real-time fashion to provide an interactive feel for the user.

## SUMMARY OF THE INVENTION

An object of the present invention, therefore, is to provide an improved system and method for overcoming the limitations of the prior art noted above;

6

A primary object of the present invention is to provide a word and phrase recognition system that is flexibly and optimally distributed across a client/platform computing architecture, so that improved accuracy, speed and uniformity can be achieved for a wide group of users;

A further object of the present invention is to provide a speech recognition system that efficiently integrates a distributed word recognition system with a natural language processing system, so that both individual words and entire speech utterances can be quickly and accurately recognized in any number of possible languages;

A related object of the present invention is to provide an efficient query response system so that an extremely accurate, real-time set of appropriate answers can be given in response to speech-based queries;

Yet another object of the present invention is to provide an interactive, real-time instructional/learning system that is distributed across a client/server architecture, and permits a real-time question/answer session with an interactive character;

A related object of the present invention is to implement such interactive character with an articulated response capability so that the user experiences a human-like interaction;

Still a further object of the present invention is to provide an INTERNET website with speech processing capability so that voice based data and commands can be used to interact with such site, thus enabling voice-based e-commerce and e-support services to be easily scaleable;

Another object is to implement a distributed speech recognition system that utilizes environmental variables as part of the recognition process to improve accuracy and speed;

A further object is to provide a scaleable query/response database system, to support any number of query topics and users as needed for a particular application and instantaneous demand;

Yet another object of the present invention is to provide a query recognition system that employs a two-step approach, including a relatively rapid first step to narrow down the list of potential responses to a smaller candidate set, and a second more computationally intensive second step to identify the best choice to be returned in response to the query from the candidate set;

A further object of the present invention is to provide a natural language processing system that facilitates query recognition by extracting lexical components of speech utterances, which components can be used for rapidly identifying a candidate set of potential responses appropriate for such speech utterances;

Another related object of the present invention is to provide a natural language processing system that facilitates query recognition by comparing lexical components of speech utterances with a candidate set of potential response to provide an extremely accurate best response to such query.

One general aspect of the present invention, therefore, relates to a natural language query system (NLQS) that offers a fully interactive method for answering user's questions over a distributed network such as the INTERNET or a local intranet. This interactive system when implemented over the worldwide web (WWW) services of the INTERNET functions so that a client or user can ask a question in a natural language such as English, French, German or Spanish and receive the appropriate answer at his or her personal computer also in his or her native natural language.

The system is distributed and consists of a set of integrated software modules at the client's machine and another

US 7,050,977 B1

7

set of integrated software programs resident on a server or set of servers. The client-side software program is comprised of a speech recognition program, an agent and its control program, and a communication program. The server-side program is comprised of a communication program, a natural language engine (NLE), a database processor (DBProcess), an interface program for interfacing the DBProcess with the NLE, and a SQL database. In addition, the client's machine is equipped with a microphone and a speaker. Processing of the speech utterance is divided between the client and server side so as to optimize processing and transmission latencies, and so as to provide support for even very thin client platforms.

In the context of an interactive learning application, the system is specifically used to provide a single-best answer to a user's question. The question that is asked at the client's machine is articulated by the speaker and captured by a microphone that is built in as in the case of a notebook computer or is supplied as a standard peripheral attachment. Once the question is captured, the question is processed partially by NLQS client-side software resident in the client's machine. The output of this partial processing is a set of speech vectors that are transported to the server via the INTERNET to complete the recognition of the user's questions. This recognized speech is then converted to text at the server.

After the user's question is decoded by the speech recognition engine (SRE) located at the server, the question is converted to a structured query language (SQL) query. This query is then simultaneously presented to a software process within the server called DBProcess for preliminary processing and to a Natural Language Engine (NLE) module for extracting the noun phrases (NP) of the user's question. During the process of extracting the noun phrase within the NLE, the tokens of the users' question are tagged. The tagged tokens are then grouped so that the NP list can be determined. This information is stored and sent to the DBProcess process.

In the DBProcess, the SQL query is fully customized using the NP extracted from the user's question and other environment variables that are relevant to the application. For example, in a training application, the user's selection of course, chapter and or section would constitute the environment variables. The SQL query is constructed using the extended SQL Full-Text predicates —CONTAINS, FREE-TEXT, NEAR, AND. The SQL query is next sent to the Full-Text search engine within the SQL database, where a Full-Text search procedure is initiated. The result of this search procedure is recordset of answers. This recordset contains stored questions that are similar linguistically to the user's question. Each of these stored questions has a paired answer stored in a separate text file, whose path is stored in a table of the database.

The entire recordset of returned stored answers is then returned to the NLE engine in the form of an array. Each stored question of the array is then linguistically processed sequentially one by one. This linguistic processing constitutes the second step of a 2-step algorithm to determine the single best answer to the user's question. This second step proceeds as follows: for each stored question that is returned in the recordset, a NP of the stored question is compared with the NP of the user's question. After all stored questions of the array are compared with the user's question, the stored question that yields the maximum match with the user's question is selected as the best possible stored question that

8

matches the user's question. The metric that is used to determine the best possible stored question is the number of noun phrases.

The stored answer that is paired to the best-stored question is selected as the one that answers the user's question. The ID tag of the question is then passed to the DBProcess. This DBProcess returns the answer which is stored in a file.

A communication link is again established to send the answer back to the client in compressed form. The answer once received by the client is decompressed and articulated to the user by the text-to-speech engine. Thus, the invention can be used in any number of different applications involving interactive learning systems, INTERNET related commerce sites, INTERNET search engines, etc.

Computer-assisted instruction environments often require the assistance of mentors or live teachers to answer questions from students. This assistance often takes the form of organizing a separate pre-arranged forum or meeting time that is set aside for chat sessions or live call-in sessions so that at a scheduled time answers to questions may be provided. Because of the time immediacy and the on-demand or asynchronous nature of on-line training where a student may log on and take instruction at any time and at any location, it is important that answers to questions be provided in a timely and cost-effective manner so that the user or student can derive the maximum benefit from the material presented.

This invention addresses the above issues. It provides the user or student with answers to questions that are normally channeled to a live teacher or mentor. This invention provides a single-best answer to questions asked by the student. The student asks the question in his or her own voice in the language of choice. The speech is recognized and the answer to the question is found using a number of technologies including distributed speech recognition, full-text search database processing, natural language processing and text-to-speech technologies. The answer is presented to the user, as in the case of a live teacher, in an articulated manner by an agent that mimics the mentor or teacher, and in the language of choice—English, French, German, Japanese or other natural spoken language. The user can choose the agent's gender as well as several speech parameters such as pitch, volume and speed of the character's voice.

Other applications that benefit from NLQS are e-commerce applications. In this application, the user's query for a price of a book, compact disk or for the availability of any item that is to be purchased can be retrieved without the need to pick through various lists on successive web pages. Instead, the answer is provided directly to the user without any additional user input.

Similarly, it is envisioned that this system can be used to provide answers to frequently-asked questions (FAQs), and as a diagnostic service tool for e-support. These questions are typical of a give web site and are provided to help the user find information related to a payment procedure or the specifications of, or problems experienced with a product/service. In all of these applications, the NLQS architecture can be applied.

A number of inventive methods associated with these architectures are also beneficially used in a variety of INTERNET related applications.

Although the inventions are described below in a set of preferred embodiments, it will be apparent to those skilled in the art the present inventions could be beneficially used in many environments where it is necessary to implement fast, accurate speech recognition, and/or to provide a human-like dialog capability to an intelligent system.

## US 7,050,977 B1

9

## BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a preferred embodiment of a natural language query system (NLQS) of the present invention, which is distributed across a client/server computing architecture, and can be used as an interactive learning system, an e-commerce system, an e-support system, and the like;

FIG. 2 is a block diagram of a preferred embodiment of a client side system, including speech capturing modules, partial speech processing modules, encoding modules, transmission modules, agent control modules, and answer/voice feedback modules that can be used in the aforementioned NLQS;

FIG. 2—2 is a block diagram of a preferred embodiment of a set of initialization routines and procedures used for the client side system of FIG. 2;

FIG. 3 is a block diagram of a preferred embodiment of a set of routines and procedures used for handling an iterated set of speech utterances on the client side system of FIG. 2, transmitting speech data for such utterances to a remote server, and receiving appropriate responses back from such server;

FIG. 4 is a block diagram of a preferred embodiment of a set of initialization routines and procedures used for un-initializing the client side system of FIG. 2;

FIG. 4A is a block diagram of a preferred embodiment of a set of routines and procedures used for implementing a distributed component of a speech recognition module for the server side system of FIG. 5;

FIG. 4B is a block diagram of a preferred set of routines and procedures used for implementing an SQL query builder for the server side system of FIG. 5;

FIG. 4C is a block diagram of a preferred embodiment of a set of routines and procedures used for implementing a database control process module for the server side system of FIG. 5;

FIG. 4D is a block diagram of a preferred embodiment of a set of routines and procedures used for implementing a natural language engine that provides query formulation support, a query response module, and an interface to the database control process module for the server side system of FIG. 5;

FIG. 5 is a block diagram of a preferred embodiment of a server side system, including a speech recognition module to complete processing of the speech utterances, environmental and grammar control modules, query formulation modules, a natural language engine, a database control module, and a query response module that can be used in the aforementioned NLQS;

FIG. 6 illustrates the organization of a full-text database used as part of server side system shown in FIG. 5;

FIG. 7A illustrates the organization of a full-text database course table used as part of server side system shown in FIG. 5 for an interactive learning embodiment of the present invention;

FIG. 7B illustrates the organization of a full-text database chapter table used as part of server side system shown in FIG. 5 for an interactive learning embodiment of the present invention;

FIG. 7C describes the fields used in a chapter table used as part of server side system shown in FIG. 5 for an interactive learning embodiment of the present invention;

FIG. 7D describes the fields used in a section table used as part of server side system shown in FIG. 5 for an interactive learning embodiment of the present invention;

10

FIG. 8 is a flow diagram of a first set of operations performed by a preferred embodiment of a natural language engine on a speech utterance including Tokenization, Tagging and Grouping;

FIG. 9 is a flow diagram of the operations performed by a preferred embodiment of a natural language engine on a speech utterance including stemming and Lexical Analysis

FIG. 10 is a block diagram of a preferred embodiment of a SQL database search and support system for the present invention;

FIGS. 11A–11C are flow diagrams illustrating steps performed in a preferred two step process implemented for query recognition by the NLQS of FIG. 2;

FIG. 12 is an illustration of another embodiment of the present invention implemented as part of a Web-based speech based learning/training System;

FIGS. 13–17 are illustrations of another embodiment of the present invention implemented as part of a Web-based e-commerce system;

FIG. 18 is an illustration of another embodiment of the present invention implemented as part of a voice-based Help Page for an E-Commerce Web Site.

## DETAILED DESCRIPTION OF THE INVENTION

## Overview

As alluded to above, the present inventions allow a user to ask a question in a natural language such as English, French, German, Spanish or Japanese at a client computing system (which can be as simple as a personal digital assistant or cell-phone, or as sophisticated as a high end desktop PC) and receive an appropriate answer from a remote server also in his or her native natural language. As such, the embodiment of the invention shown in FIG. 1 is beneficially used in what can be generally described as a Natural Language Query System (NLQS) 100, which is configured to interact on a real-time basis to give a human-like dialog capability/experience for e-commerce, e-support, and e-learning applications.

The processing for NLQS 100 is generally distributed across a client side system 150, a data link 160, and a server-side system 180. These components are well known in the art, and in a preferred embodiment include a personal computer system 150, an INTERNET connection 160A, 160B, and a larger scale computing system 180. It will be understood by those skilled in the art that these are merely exemplary components, and that the present invention is by no means limited to any particular implementation or combination of such systems. For example, client-side system 150 could also be implemented as a computer peripheral, a PDA, as part of a cell-phone, as part of an INTERNET-adapted appliance, an INTERNET linked kiosk, etc. Similarly, while an INTERNET connection is depicted for data link 160A, it is apparent that any channel that is suitable for carrying data between client system 150 and server system 180 will suffice, including a wireless link, an RF link, an IR link, a LAN, and the like. Finally, it will be further appreciated that server system 180 may be a single, large-scale system, or a collection of smaller systems interlinked to support a number of potential network users.

Initially speech input is provided in the form of a question or query articulated by the speaker at the client's machine or personal accessory as a speech utterance. This speech utterance is captured and partially processed by NLQS client-side software 155 resident in the client's machine. To



US 7,050,977 B1

11

facilitate and enhance the human-like aspects of the interaction, the question is presented in the presence of an animated character **157** visible to the user who assists the user as a personal information retriever/agent. The agent can also interact with the user using both visible text output on a monitor/display (not shown) and/or in audible form using a text to speech engine **159**. The output of the partial processing done by SRE **155** is a set of speech vectors that are transmitted over communication channel **160** that links the user's machine or personal accessory to a server or servers via the INTERNET or a wireless gateway that is linked to the INTERNET as explained above. At server **180**, the partially processed speech signal data is handled by a server-side SRE **182**, which then outputs recognized speech text corresponding to the user's question. Based on this user question related text, a text-to-query converter **184** formulates a suitable query that is used as input to a database processor **186**. Based on the query, database processor **186** then locates and retrieves an appropriate answer using a customized SQL query from database **188**. A Natural Language Engine **190** facilitates structuring the query to database **188**. After a matching answer to the user's question is found, the former is transmitted in text form across data link **160B**, where it is converted into speech by text to speech engine **159**, and thus expressed as oral feedback by animated character agent **157**.

Because the speech processing is broken up in this fashion, it is possible to achieve real-time, interactive, human-like dialog consisting of a large, controllable set of questions/answers. The assistance of the animated agent **157** further enhances the experience, making it more natural and comfortable for even novice users. To make the speech recognition process more reliable, context-specific grammars and dictionaries are used, as well as natural language processing routines at NLE **190**, to analyze user questions lexically. While context-specific processing of speech data is known in the art (see e.g., U.S. Pat. Nos. 5,960,394, 5,867, 817, 5,758,322 and 5,384,892 incorporated by reference herein) the present inventors are unaware of any such implementation as embodied in the present inventions. The text of the user's question is compared against text of other questions to identify the question posed by the user by DB processor/engine (DBE) **186**. By optimizing the interaction and relationship of the SR engines **155** and **182**, the NLP routines **190**, and the dictionaries and grammars, an extremely fast and accurate match can be made, so that a unique and responsive answer can be provided to the user.

On the server side **180**, interleaved processing further accelerates the speech recognition process. In simplified terms, the query is presented simultaneously both to NLE **190** after the query is formulated, as well as to DBE **186**. NLE **190** and SRE **182** perform complementary functions in the overall recognition process. In general, SRE **182** is primarily responsible for determining the identity of the words articulated by the user, while NLE **190** is responsible for the linguistic morphological analysis of both the user's query and the search results returned after the database query.

After the user's query is analyzed by NLE **190** some parameters are extracted and sent to the DBProcess. Additional statistics are stored in an array for the 2<sup>nd</sup> step of processing. During the 2<sup>nd</sup> step of 2-step algorithm, the recordset of preliminary search results are sent to the NLE **160** for processing. At the end of this 2<sup>nd</sup> step, the single question that matches the user's query is sent to the DBProcess where further processing yields the paired answer that is paired with the single best stored question.

12

Thus, the present invention uses a form of natural language processing (NLP) to achieve optimal performance in a speech based web application system. While NLP is known in the art, prior efforts in Natural Language Processing (NLP) work nonetheless have not been well integrated with Speech Recognition (SR) technologies to achieve reasonable results in a web-based application environment. In speech recognition, the result is typically a lattice of possible recognized words each with some probability of fit with the speech recognizer. As described before, the input to a typical NLP system is typically a large linguistic unit. The NLP system is then charged with the parsing, understanding and indexing of this large linguistic unit or set of transcribed utterances. The result of this NLP process is to understand lexically or morphologically the entire linguistic unit as opposed to word recognition. Put another way, the linguistic unit or sentence of connected words output by the SRE has to be understood lexically, as opposed to just being "recognized".

As indicated earlier, although speech recognition technology has been available for several years, the technical requirements for the NLQS invention have placed severe restrictions on the specifications for the speech recognition accuracy that is required for an application that combines speech recognition and natural language processing to work satisfactorily. In realizing that even with the best of conditions, it might be not be possible to achieve the perfect 100% speech recognition accuracy that is required, the present invention employs an algorithm that balances the potential risk of the speech recognition process with the requirements of the natural language processing so that even in cases where perfect speech recognition accuracy is not achieved for each word in the query, the entire query itself is nonetheless recognized with sufficient accuracy.

This recognition accuracy is achieved even while meeting very stringent user constraints, such as short latency periods of 3 to 5 seconds (ideally—ignoring transmission latencies which can vary) for responding to a speech-based query, and for a potential set of 100–250 query questions. This quick response time gives the overall appearance and experience of a real-time discourse that is more natural and pleasant from the user's perspective. Of course, non-real time applications, such as translation services for example, can also benefit from the present teachings as well, since a centralized set of HMMs, grammars, dictionaries, etc., are maintained.

#### General Aspects of Speech Recognition Used in the Present Inventions

General background information on speech recognition can be found in the prior art references discussed above and incorporated by reference herein. Nonetheless, a discussion of some particular exemplary forms of speech recognition structures and techniques that are well-suited for NLQS **100** is provided next to better illustrate some of the characteristics, qualities and features of the present inventions.

Speech recognition technology is typically of two types—speaker independent and speaker dependent. In speaker-dependent speech recognition technology, each user has a voice file in which a sample of each potentially recognized word is stored. Speaker-dependent speech recognition systems typically have large vocabularies and dictionaries making them suitable for applications as dictation and text transcribing. It follows also that the memory and processor resource requirements for the speaker-dependent can be and are typically large and intensive.

US 7,050,977 B1

13

Conversely speaker-independent speech recognition technology allows a large group of users to use a single vocabulary file. It follows then that the degree of accuracy that can be achieved is a function of the size and complexity of the grammars and dictionaries that can be supported for a given language. Given the context of applications for which NLQS, the use of small grammars and dictionaries allow speaker independent speech recognition technology to be implemented in NLQS.

The key issues or requirements for either type—speaker-independent or speaker-dependent, are accuracy and speed. As the size of the user dictionaries increase, the speech recognition accuracy metric—word error rate (WER) and the speed of recognition decreases. This is so because the search time increases and the pronunciation match becomes more complex as the size of the dictionary increases.

The basis of the NLQS speech recognition system is a series of Hidden Markov Models (HMM), which, as alluded to earlier, are mathematical models used to characterize any time varying signal. Because parts of speech are considered to be based on an underlying sequence of one or more symbols, the HMM models corresponding to each symbol are trained on vectors from the speech waveforms. The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multi-dimensional) probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to an associated probability distribution. This finite state machine changes state once every time unit, and each time  $t$  such that a state  $j$  is entered, a spectral parameter vector  $O_t$  is generated with probability density  $B_j(O_t)$ . It is only the outcome, not the state which is visible to an external observer and therefore states are “hidden” to the outside; hence the name Hidden Markov Model.

In isolated speech recognition, it is assumed that the sequence of observed speech vectors corresponding to each word can each be described by a Markov model as follows:

$$O=O_1, O_2, \dots, O_T \quad (1-1)$$

where  $O_t$  is a speech vector observed at time  $t$ . The isolated word recognition then is to compute:

$$\arg \max \{P(w_i|O)\} \quad (1-2)$$

By using Bayes' Rule,

$$\{P(w_i|O)\}=[P(O|w_i)P(w_i)]/P(O) \quad (1-3)$$

In the general case, the Markov model when applied to speech also assumes a finite state machine which changes state once every time unit and each time that a state  $j$  is entered, a speech vector  $O_t$  is generated from the probability density  $b_j(O_t)$ . Furthermore, the transition from state  $i$  to state  $j$  is also probabilistic and is governed by the discrete probability  $a_{ij}$ .

For a state sequence  $X$ , the joint probability that  $O$  is generated by the model  $M$  moving through a state sequence  $X$  is the product of the transition probabilities and the output probabilities. Only the observation sequence is known—the state sequence is hidden as mentioned before.

Given that  $X$  is unknown, the required likelihood is computed by summing over all possible state sequences  $X=x(1), x(2), x(3), \dots, x(T)$ , that is

$$P(O|M)=\sum_{x(1), x(2), \dots, x(T)} \prod_{t=1}^T b(x(t))a_{x(t)x(t+1)}$$

Given a set of models  $M_i$ , corresponding to words  $w_i$  equation 1-2 is solved by using 1-3 and also by assuming that:

14

$$P(O|w_i)=P(O|M_i)$$

All of this assumes that the parameters  $\{a_{ij}\}$  and  $\{b_j(O_t)\}$  are known for each model  $M_i$ . This can be done, as explained earlier, by using a set of training examples corresponding to a particular model. Thereafter, the parameters of that model can be determined automatically by a robust and efficient re-estimation procedure. So if a sufficient number of representative examples of each word are collected, then a HMM can be constructed which simply models all of the many sources of variability inherent in real speech. This training is well-known in the art, so it is not described at length herein, except to note that the distributed architecture of the present invention enhances the quality of HMMs, since they are derived and constituted at the server side, rather than the client side. In this way, appropriate samples from users of different geographical areas can be easily compiled and analyzed to optimize the possible variations expected to be seen across a particular language to be recognized. Uniformity of the speech recognition process is also well-maintained, and error diagnostics are simplified, since each prospective user is using the same set of HMMs during the recognition process.

To determine the parameters of a HMM from a set of training samples, the first step typically is to make a rough guess as to what they might be. Then a refinement is done using the Baum-Welch estimation formulae. By these formulae, the maximum likelihood estimates of  $\mu_j$  (where  $\mu_j$  is mean vector and  $\Sigma_j$  is covariance matrix) is:

$$\mu_j=\Sigma_{t=1}^T L_j(t)O_t / [\Sigma_{t=1}^T L_j(t)]$$

A forward-backward algorithm is next used to calculate the probability of state occupation  $L_j(t)$ . If the forward probability  $\alpha_j(t)$  for some model  $M$  with  $N$  states is defined as:

$$\alpha_j(t)=P(O_1, \dots, O_t, x(t)=j|M)$$

This probability can be calculated using the recursion:

$$\alpha_j(t)=[\Sigma_{i=1}^{N-1} \alpha_i(t-1)a_{ij}]b_j(O_t)$$

Similarly the backward probability can be computed using the recursion:

$$\beta_j(t)=\Sigma_{i=2}^{N-1} a_{ij}b_j(O_{t+1})\beta_i(t+1)$$

Realizing that the forward probability is a joint probability and the backward probability is a conditional probability, the probability of state occupation is the product of the two probabilities:

$$\alpha_j(t)\beta_j(t)=P(O, x(t)=j|M)$$

Hence the probability of being in state  $j$  at a time  $t$  is:

$$L_j(t)=1/P[\alpha_j(t)\beta_j(t)]$$

where  $P=P(O|M)$

To generalize the above for continuous speech recognition, we assume the maximum likelihood state sequence where the summation is replaced by a maximum operation. Thus for a given model  $M$ , let  $\phi_j(t)$  represent the maximum likelihood of observing speech vectors  $O_1$  to  $O_t$  and being used in state  $j$  at time  $t$ :

$$\phi_j(t)=\max \{\phi_j(t-1)\alpha_{ij}\}b_j(O_t)$$

Expressing this logarithmically to avoid underflow, this likelihood becomes:

$$\psi_j(t)=\max \{\psi_j(t-1)+\log(\alpha_{ij})\}+\log(b_j(O_t))$$

This is also known as the Viterbi algorithm. It can be visualized as finding the best path through a matrix where the vertical dimension represents the states of the HMM and

US 7,050,977 B1

15

horizontal dimension represents frames of speech i.e. time. To complete the extension to connected speech recognition, it is further assumed that each HMM representing the underlying sequence is connected. Thus the training data for continuous speech recognition should consist of connected utterances; however, the boundaries between words do not have to be known.

To improve computational speed/efficiency, the Viterbi algorithm is sometimes extended to achieve convergence by using what is known as a Token Passing Model. The token passing model represents a partial match between the observation sequence  $o_1$  to  $o_t$  and a particular model, subject to the constraint that the model is in state  $j$  at time  $t$ . This token passing model can be extended easily to connected speech environments as well if we allow the sequence of HMMs to be defined as a finite state network. A composite network that includes both phoneme-based HMMs and complete words can be constructed so that a single-best word can be recognized to form connected speech using word N-best extraction from the lattice of possibilities. This composite form of HMM-based connected speech recognizer is the basis of the NLQS speech recognizer module. Nonetheless, the present invention is not limited as such to such specific forms of speech recognizers, and can employ other techniques for speech recognition if they are otherwise compatible with the present architecture and meet necessary performance criteria for accuracy and speed to provide a real-time dialog experience for users.

The representation of speech for the present invention's HMM-based speech recognition system assumes that speech is essentially either a quasi-periodic pulse train (for voiced speech sounds) or a random noise source (for unvoiced sounds). It may be modeled as two sources—one an impulse train generator with pitch period  $P$  and a random noise generator which is controlled by a voice/unvoiced switch. The output of the switch is then fed into a gain function estimated from the speech signal and scaled to feed a digital filter  $H(z)$  controlled by the vocal tract parameter characteristics of the speech being produced. All of the parameters for this model—the voiced/unvoiced switching, the pitch period for voiced sounds, the gain parameter for the speech signal and the coefficient of the digital filter, vary slowly with time. In extracting the acoustic parameters from the user's speech input so that it can be evaluated in light of a set of HMMs, cepstral analysis is typically used to separate the vocal tract information from the excitation information. The cepstrum of a signal is computed by taking the Fourier (or similar) transform of the log spectrum. The principal advantage of extracting cepstral coefficients is that they are de-correlated and the diagonal covariances can be used with HMMs. Since the human ear resolves frequencies non-linearly across the audio spectrum, it has been shown that a front-end that operates in a similar non-linear way improves speech recognition performance.

Accordingly, instead of a typical linear prediction-based analysis, the front-end of the NLQS speech recognition engine implements a simple, fast Fourier transform based filter bank designed to give approximately equal resolution on the Mel-scale. To implement this filter bank, a window of speech data (for a particular time frame) is transformed using a software based Fourier transform and the magnitude taken. Each FFT magnitude is then multiplied by the corresponding filter gain and the results accumulated. The cepstral coefficients that are derived from this filter-bank analysis at the front end are calculated during a first partial processing phase of the speech signal by using a Discrete Cosine Transform of the log filter bank amplitudes. These

16

cepstral coefficients are called Mel-Frequency Cepstral Coefficients (MFCC) and they represent some of the speech parameters transferred from the client side to characterize the acoustic features of the user's speech signal. These parameters are chosen for a number of reasons, including the fact that they can be quickly and consistently derived even across systems of disparate capabilities (i.e., for everything from a low power PDA to a high powered desktop system), they give good discrimination, they lend themselves to a number of useful recognition related manipulations, and they are relatively small and compact in size so that they can be transported rapidly across even a relatively narrow band link. Thus, these parameters represent the least amount of information that can be used by a subsequent server side system to adequately and quickly complete the recognition process.

To augment the speech parameters an energy term in the form of the logarithm of the signal energy is added. Accordingly, RMS energy is added to the 12 MFCC's to make 13 coefficients. These coefficients together make up the partially processed speech data transmitted in compressed form from the user's client system to the remote server side.

The performance of the present speech recognition system is enhanced significantly by computing and adding time derivatives to the basic static MFCC parameters at the server side. These two other sets of coefficients—the delta and acceleration coefficients representing change in each of the 13 values from frame to frame (actually measured across several frames), are computed during a second partial speech signal processing phase to complete the initial processing of the speech signal, and are added to the original set of coefficients after the latter are received. These MFCCs together with the delta and acceleration coefficients constitute the observation vector  $O_t$  mentioned above that is used for determining the appropriate HMM for the speech data.

The delta and acceleration coefficients are computed using the following regression formula:

$$d_t = \sum_{\theta=1}^{\theta} [c_{t+\theta} - c_{t-\theta}] / 2 \sum_{\theta=1}^{\theta} \theta^2$$

where  $d_t$  is a delta coefficient at time  $t$  computed in terms of the corresponding static coefficients:

$$d_t = [c_{t+\theta} - c_{t-\theta}] / 2\theta$$

In a typical stand-alone implementation of a speech recognition system, the entire SR engine runs on a single client. In other words, both the first and second partial processing phases above are executed by the same DSP (or microprocessor) running a ROM or software code routine at the client's computing machine.

In contrast, because of several considerations, specifically—cost, technical performance, and client hardware uniformity, the present NLQS system uses a partitioned or distributed approach. While some processing occurs on the client side, the main speech recognition engine runs on a centrally located server or number of servers. More specifically, as noted earlier, capture of the speech signals, MFCC vector extraction and compression are implemented on the client's machine during a first partial processing phase. The routine is thus streamlined and simple enough to be implemented within a browser program (as a plug in module, or a downloadable applet for example) for maximum ease of use and utility. Accordingly, even very "thin" client platforms can be supported, which enables the use of the present system across a greater number of potential sites. The primary MFCCs are then transmitted to the server over the channel, which, for example, can include a dial-up INTERNET connection, a LAN connection, a wireless connection



US 7,050,977 B1

17

and the like. After decompression, the delta and acceleration coefficients are computed at the server to complete the initial speech processing phase, and the resulting observation vectors  $O_t$  are also determined.

#### General Aspects of Speech Recognition Engine

The speech recognition engine is also located on the server, and is based on a HTK-based recognition network compiled from a word-level network, a dictionary and a set of HMMs. The recognition network consists of a set of nodes connected by arcs. Each node is either a HMM model instance or a word end. Each model node is itself a network consisting of states connected by arcs. Thus when fully compiled, a speech recognition network consists of HMM states connected by transitions. For an unknown input utterance with  $T$  frames, every path from the start node to the exit node of the network passes through  $T$  HMM states. Each of these paths has log probability which is computed by summing the log probability of each individual transition in the path and the log probability of each emitting state generating the corresponding observation. The function of the Viterbi decoder is find those paths through the network which have the highest log probability. This is found using the Token Passing algorithm. In a network that has many nodes, the computation time is reduced by only allowing propagation of those tokens which will have some chance of becoming winners. This process is called pruning.

#### Natural Language Processor

In a typical natural language interface to a database, the user enters a question in his/her natural language, for example, English. The system parses it and translates it to a query language expression. The system then uses the query language expression to process the query and if the search is successful, a recordset representing the results is displayed in English either formatted as raw text or in a graphical form. For a natural language interface to work well involves a number of technical requirements.

For example, it needs to be robust—in the sentence ‘What’s the departments turnover’ it needs to decide that the word *whats=what’s=what is*. And it also has to determine that *departments=department’s*. In addition to being robust, the natural language interface has to distinguish between the several possible forms of ambiguity that may exist in the natural language—lexical, structural, reference and ellipsis ambiguity. All of these requirements, in addition to the general ability to perform basic linguistic morphological operations of tokenization, tagging and grouping, are implemented within the present invention.

Tokenization is implemented by a text analyzer which treats the text as a series of tokens or useful meaningful units that are larger than individual characters, but smaller than phrases and sentences. These include words, separable parts of words, and punctuation. Each token is associated with an offset and a length. The first phase of tokenization is the process of segmentation which extracts the individual tokens from the input text and keeps track of the offset where each token originated in the input text. The tokenizer output lists the offset and category for each token. In the next phase of the text analysis, the tagger uses a built-in morphological analyzer to look up each word/token in a phrase or sentence and internally lists all parts of speech. The output is the input string with each token tagged with a parts of speech notation. Finally the grouper which functions as a phrase extractor or phrase analyzer, determines which groups of words form phrases. These three operations which are the foundations for any modern linguistic processing schemes, are fully

18

implemented in optimized algorithms for determining the single-best possible answer to the user’s question.

#### SQL Database and Full-Text Query

Another key component of present system is a SQL-database. This database is used to store text, specifically the answer-question pairs are stored in full-text tables of the database. Additionally, the full-text search capability of the database allows full-text searches to be carried out.

While a large portion of all digitally stored information is in the form of unstructured data, primarily text, it is now possible to store this textual data in traditional database systems in character-based columns such as varchar and text. In order to effectively retrieve textual data from the database, techniques have to be implemented to issue queries against textual data and to retrieve the answers in a meaningful way where it provides the answers as in the case of the NLQS system.

There are two major types of textual searches: Property—This search technology first applies filters to documents in order to extract properties such as author, subject, type, word count, printed page count, and time last written, and then issues searches against those properties; Full-text—this search technology first creates indexes of all non-noise words in the documents, and then uses these indexes to support linguistic searches and proximity searches.

Two additional technologies are also implemented in this particular RDBMs: SQL Server also have been integrated: A Search service—a full-text indexing and search service that is called both index engine and search, and a parser that accepts full-text SQL extensions and maps them into a form that can be processed by the search engine.

The four major aspects involved in implementing full-text retrieval of plain-text data from a full-text-capable database are: Managing the definition of the tables and columns that are registered for full-text searches; Indexing the data in registered columns—the indexing process scans the character streams, determines the word boundaries (this is called word breaking), removes all noise words (this also is called stop words), and then populates a full-text index with the remaining words; Issuing queries against registered columns for populated full-text indexes; Ensuring that subsequent changes to the data in registered columns gets propagated to the index engine to keep the full-text indexes synchronized.

The underlying design principle for the indexing, querying, and synchronizing processes is the presence of a full-text unique key column (or single-column primary key) on all tables registered for full-text searches. The full-text index contains an entry for the non-noise words in each row together with the value of the key column for each row.

When processing a full-text search, the search engine returns to the database the key values of the rows that match the search criteria.

The full-text administration process starts by designating a table and its columns of interest for full-text search. Customized NLQS stored procedures are used first to register tables and columns as eligible for full-text search. After that, a separate request by means of a stored procedure is issued to populate the full-text indexes. The result is that the underlying index engine gets invoked and asynchronous index population begins. Full-text indexing tracks which significant words are used and where they are located. For example, a full-text index might indicate that the word “NLQS” is found at word number 423 and word number 982 in the Abstract column of the DevTools table for the row associated with a ProductID of 6. This index structure supports an efficient search for all items containing indexed

US 7,050,977 B1

19

words as well as advanced search operations, such as phrase searches and proximity searches. (An example of a phrase search is looking for “white elephant,” where “white” is followed by “elephant”. An example of a proximity search is looking for “big” and “house” where “big” occurs near “house”.) To prevent the full-text index from becoming bloated, noise words such as “a,” “and,” and “the” are ignored.

Extensions to the Transact-SQL language are used to construct full-text queries. The two key predicates that are used in the NLQS are CONTAINS and FREETEXT.

The CONTAINS predicate is used to determine whether or not values in full-text registered columns contain certain words and phrases. Specifically, this predicate is used to search for:

A word or phrase.

The prefix of a word or phrase.

A word or phrase that is near another.

A word that is an inflectional form of another (for example, “drive” is the inflectional stem of “drives,” “drove,” “driving,” and “driven”).

A set of words or phrases, each of which is assigned a different weighting.

The relational engine within SQL Server recognizes the CONTAINS and FREETEXT predicates and performs some minimal syntax and semantic checking, such as ensuring that the column referenced in the predicate has been registered for full-text searches. During query execution, a full-text predicate and other relevant information are passed to the full-text search component. After further syntax and semantic validation, the search engine is invoked and returns the set of unique key values identifying those rows in the table that satisfy the full-text search condition. In addition to the FREETEXT and CONTAINS, other predicates such as AND, LIKE, NEAR are combined to create the customized NLQS SQL construct.

#### Full-Text Query Architecture of the SQL Database

The full-text query architecture is comprised of the following several components—Full-Text Query component, the SQL Server Relational Engine, the Full-Text provider and the Search Engine.

The Full-Text Query component of the SQL database accept a full-text predicate or rowset-valued function from the SQL Server; transform parts of the predicate into an internal format, and sends it to Search Service, which returns the matches in a rowset. The rowset is then sent back to SQL Server. SQL Server uses this information to create the resultset that is then returned to the submitter of the query.

The SQL Server Relational Engine accepts the CONTAINS and FREETEXT predicates as well as the CONTAINSTABLE() and FREETEXTTABLE() rowset-valued functions. During parse time, this code checks for conditions such as attempting to query a column that has not been registered for full-text search. If valid, then at run time, the ft\_search\_condition and context information is sent to the full-text provider. Eventually, the full-text provider returns a rowset to SQL Server, which is used in any joins (specified or implied) in the original query. The Full-Text Provider parses and validates the ft\_search\_condition, constructs the appropriate internal representation of the full-text search condition, and then passes it to the search engine. The result is returned to the relational engine by means of a rowset of rows that satisfy ft\_search\_condition.

#### Client Side System 150

The architecture of client-side system 150 of Natural Language Query System 100 is illustrated in greater detail in

20

FIG. 2. Referring to FIG. 2, the three main processes effectuated by Client System 150 are illustrated as follows: Initialization process 200A consisting of SRE 201, Communication 202 and Microsoft (MS) Agent 203 routines; an iterative process 200B consisting of two sub-routines: a) Receive User Speech 208—made up of SRE 204 and Communication 205; and b) Receive Answer from Server 207—made up of MS Speak Agent 206, Communication 209, Voice data file 210 and Text to Speech Engine 211. Finally, un-initialization process 200C is made up of three sub-routines: SRE 212, Communication 213, and MS Agent 214. Each of the above three processes are described in detail in the following paragraphs. It will be appreciated by those skilled in the art that the particular implementation for such processes and routines will vary from client platform to platform, so that in some environments such processes may be embodied in hard-coded routines executed by a dedicated DSP, while in others they may be embodied as software routines executed by a shared host processor, and in still others a combination of the two may be used.

#### Initialization at Client System 150

The initialization of the Client System 150 is illustrated in FIG. 2—2 and is comprised generally of 3 separate initializing processes: client-side Speech Recognition Engine 220A, MS Agent 220B and Communication processes 220C.

#### Initialization of Speech Recognition Engine 220A

Speech Recognition Engine 155 is initialized and configured using the routines shown in 220A. First, an SRE COM Library is initialized. Next, memory 220 is allocated to hold Source and Coder objects, are created by a routine 221. Loading of configuration file 221A from configuration data file 221B also takes place at the same time that the SRE Library is initialized. In configuration file 221B, the type of the input of Coder and the type of the output of the Coder are declared. The structure, operation, etc. of such routines are well-known in the art, and they can be implemented using a number of fairly straightforward approaches. Accordingly, they are not discussed in detail herein. Next, Speech and Silence components of an utterance are calibrated using a routine 222, in a procedure that is also well-known in the art. To calibrate the speech and silence components, the user preferably articulates a sentence that is displayed in a text box on the screen. The SRE library then estimates the noise and other parameters required to find e silence and speech elements of future user utterances.

#### Initialization of MS Agent 220B

The software code used to initialize and set up a MS Agent 220B is also illustrated in FIG. 2—2. The MS Agent 220B routine is responsible for coordinating and handling the actions of the animated agent 157 (FIG. 1). This initialization thus consists of the following steps:

1. Initialize COM library 223. This part of the code initializes the COM library, which is required to use ActiveX Controls, which controls are well-known in the art.
2. Create instance of Agent Server 224—this part of the code creates an instance of Agent ActiveX control.
3. Loading of MS Agent 225—this part of the code loads MS Agent character from a specified file 225A containing general parameter data for the Agent Character, such as the overall appearance, shape, size, etc.
4. Get Character Interface 226—this part of the code gets an appropriate interface for the specified character; for

US 7,050,977 B1

21

example, characters may have different control/interaction capabilities that can be presented to the user.

5. Add Commands to Agent Character Option **227**—this part of the code adds commands to an Agent Properties sheet, which sheet can be accessed by clicking on the icon that appears in the system tray, when the Agent character is loaded e.g., that the character can Speak, how he/she moves, TTS Properties, etc.
6. Show the Agent Character **228**—this part of the code displays the Agent character on the screen so it can be seen by the user;
7. AgentNotifySink—to handle events. This part of the code creates AgentNotifySink object **229**, registers it at **230** and then gets the Agent Properties interface **231**. The property sheet for the Agent character is assigned using routine **232**.
8. Do Character Animations **233**—This part of the code plays specified character animations to welcome the user to NLQS **100**.

The above then constitutes the entire sequence required to initialize the MS Agent. As with the SRE routines, the MS Agent routines can be implemented in any suitable and conventional fashion by those skilled in the art based on the present teachings. The particular structure, operation, etc. of such routines is not critical, and thus they are not discussed in detail herein.

In a preferred embodiment, the MS Agent is configured to have an appearance and capabilities that are appropriate for the particular application. For instance, in a remote learning application, the agent has the visual form and mannerisms/attitude/gestures of a college professor. Other visual props (blackboard, textbook, etc.) may be used by the agent and presented to the user to bring to mind the experience of being in an actual educational environment. The characteristics of the agent may be configured at the client side **150**, and/or as part of code executed by a browser program (not shown) in response to configuration data and commands from a particular web page. For example, a particular website offering medical services may prefer to use a visual image of a doctor. These and many other variations will be apparent to those skilled in the art for enhancing the human-like, real-time dialog experience for users.

#### Initialization of Communication Link **160A**

The initialization of Communication Link **160A** is shown with reference to process **220C** FIG. 2—2. Referring to FIG. 2—2, this initialization consists of the following code components: Open INTERNET Connection **234**—this part of the code opens an INTERNET Connection and sets the parameter for the connection. Then Set Callback Status routine **235** sets the callback status so as to inform the user of the status of connection. Finally Start New HTTP INTERNET Session **236** starts a new INTERNET session. The details of Communications Link **160** and the set up process **220C** are not critical, and will vary from platform to platform. Again, in some cases, users may use a low-speed dial-up connection, a dedicated high speed switched connection (T1 for example), an always-on xDSL connection, a wireless connection, and the like.

#### Iterative Processing of Queries/Answers

As illustrated in FIG. 3, once initialization is complete, an iterative query/answer process is launched when the user presses the Start Button to initiate a query. Referring to FIG. 3, the iterative query/answer process consists of two main sub-processes implemented as routines on the client side system **150**: Receive User Speech **240** and Receive User Answer **243**. The Receive User Speech **240** routine receives

22

speech from the user (or another audio input source), while the Receive User Answer **243** routine receives an answer to the user's question in the form of text from the server so that it can be converted to speech for the user by text-to-speech engine **159**. As used herein, the term "query" is referred to in the broadest sense to refer, to either a question, a command, or some form of input used as a control variable by the system. For example, a query may consist of a question directed to a particular topic, such as "what is a network" in the context of a remote learning application. In an e-commerce application a query might consist of a command to "list all books by Mark Twain" for example. Similarly, while the answer in a remote learning application consists of text that is tendered into audible form by the text to speech engine **159**, it could also be returned as another form of multi-media information, such as a graphic image, a sound file, a video file, etc. depending on the requirements of the particular application. Again, given the present teachings concerning the necessary structure, operation, functions, performance, etc., of the client-side Receive User Speech **240** and Receiver User Answer **243** routines, one of ordinary skill in the art could implement such in a variety of ways.

Receive User Speech—As illustrated in FIG. 3, the Receive User Speech routine **240** consists of a SRE **241** and a Communication **242** process, both implemented again as routines on the client side system **150** for receiving and partially processing the user's utterance. SRE routine **241** uses a coder **248** which is prepared so that a coder object receives speech data from a source object. Next the Start Source **249** routine is initiated. This part of the code initiates data retrieval using the source Object which will in turn be given to the Coder object. Next, MFCC vectors **250** are extracted from the Speech utterance continuously until silence is detected. As alluded to earlier, this represents the first phase of processing of the input speech signal, and in a preferred embodiment, it is intentionally restricted to merely computing the MFCC vectors for the reasons already expressed above. These vectors include the 12 cepstral coefficients and the RMS energy term, for a total of 13 separate numerical values for the partially processed speech signal.

In some environments, nonetheless, it is conceivable that the MFCC delta parameters and MFCC acceleration parameters can also be computed at client side system **150**, depending on the computation resources available, the transmission bandwidth in data link **160A** available to server side system **180**, the speed of a transceiver used for carrying data in the data link, etc. These parameters can be determined automatically by client side system upon initializing SRE **155** (using some type of calibration routine to measure resources), or by direct user control, so that the partitioning of signal processing responsibilities can be optimized on a case-by-case basis. In some applications, too, server side system **180** may lack the appropriate resources or routines for completing the processing of the speech input signal. Therefore, for some applications, the allocation of signal processing responsibilities may be partitioned differently, to the point where in fact both phases of the speech signal processing may take place at client side system **150** so that the speech signal is completely—rather than partially—processed and transmitted for conversion into a query at server side system **180**.

Again in a preferred embodiment, to ensure reasonable accuracy and real-time performance from a query/response perspective, sufficient resources are made available in a client side system so that 100 frames per second of speech data can be partially processed and transmitted through link



## US 7,050,977 B1

23

160A. Since the least amount of information that is necessary to complete the speech recognition process (only 13 coefficients) is sent, the system achieves a real-time performance that is believed to be highly optimized, because other latencies (i.e., client-side computational latencies, packet formation latencies, transmission latencies) are minimized. It will be apparent that the principles of the present invention can be extended to other SR applications where some other methodology is used for breaking down the speech input signal by an SRE (i.e., non-MFCC based). The only criteria is that the SR processing be similarly dividable into multiple phases, and with the responsibility for different phases being handled on opposite sides of link 160A depending on overall system performance goals, requirements and the like. This functionality of the present invention can thus be achieved on a system-by-system basis, with an expected and typical amount of optimization being necessary for each particular implementation.

Thus, the present invention achieves a response rate performance that is tailored in accordance with the amount of information that is computed, coded and transmitted by the client side system 150. So in applications where real-time performance is most critical, the least possible amount of extracted speech data is transmitted to reduce these latencies, and, in other applications, the amount of extracted speech data that is processed, coded and transmitted can be varied.

Communication—transmit communication module 242 is used to implement the transport of data from the client to the server over the data link 160A, which in a preferred embodiment is the INTERNET. As explained above, the data consists of encoded MFCC vectors that will be used at then server-side of the Speech Recognition engine to complete the speech recognition decoding. The sequence of the communication is as follows:

OpenHTTPRequest 251—this part of the code first converts MFCC vectors to a stream of bytes, and then processes the bytes so that it is compatible with a protocol known as HTTP. This protocol is well-known in the art, and it is apparent that for other data links another suitable protocol would be used.

1. Encode MFCC Byte Stream 251—this part of the code encodes the MFCC vectors, so that they can be sent to the server via HTTP.
2. Send data 252—this part of the code sends MFCC vectors to the server using the INTERNET connection and the HTTP protocol.

Wait for the Server Response 253—this part of the code monitors the data link 160A a response from server side system 180 arrives. In summary, the MFCC parameters are extracted or observed on-the-fly from the input speech signal. They are then encoded to a HTTP byte stream and sent in a streaming fashion to the server before the silence is detected—i.e. sent to server side system 180 before the utterance is complete. This aspect of the invention also facilitates a real-time behavior, since data can be transmitted and processed even while the user is still speaking.

Receive Answer from Server 243 is comprised of the following modules as shown in FIG. 3.: MS Agent 244, Text-to-Speech Engine 245 and receive communication modules 246. All three modules interact to receive the answer from server side system 180. As illustrated in FIG. 3, the receive communication process consists of three separate processes implemented as a receive routine on client side system 150: a Receive the Best Answer 258 receives the best answer over data link 160B (the HTTP communication channel). The answer is de-compressed at

24

259 and then the answer is passed by code 260 to the MS Agent 244, where it is received by code portion 254. A routine 255 then articulates the answer using text-to-speech engine 257. Of course, the text can also be displayed for additional feedback purposes on a monitor used with client side system 150. The text to speech engine uses a natural language voice data file 256 associated with it that is appropriate for the particular language application (i.e., English, French, German, Japanese, etc.). As explained earlier when the answer is something more than text, it can be treated as desired to provide responsive information to the user, such as with a graphics image, a sound, a video clip, etc.

#### Uninitialization

The un-initialization routines and processes are illustrated in FIG. 4. Three functional modules are used for un-initializing the primary components of the client side system 150; these include SRE 270, Communications 271 and MS Agent 272 un-initializing routines. To un-initialize SRE 220A, memory that was allocated in the initialization phase is de-allocated by code 273 and objects created during such initialization phase are deleted by code 274. Similarly, as illustrated in FIG. 4, to un-initialize Communications module 220C the INTERNET connection previously established with the server is closed by code portion 275 of the Communication Un-initialization routine 271. Next the INTERNET session created at the time of initialization is also closed by routine 276. For the un-initialization of the MS Agent 220B, as illustrated in FIG. 4, MS Agent Un-initialization routine 272 first releases the Commands Interface 227 using routine 277. This releases the commands added to the property sheet during loading of the agent character by routine 225. Next the Character Interface initialized by routine 226 is released by routine 278 and the Agent is unloaded at 279. The Sink Object Interface is then also released 280 followed by the release of the Property Sheet Interface 281. The Agent Notify Sink 282 then un-registers the Agent and finally the Agent Interface 283 is released which releases all the resources allocated during initialization steps identified in FIG. 2—2.

It will be appreciated by those skilled in the art that the particular implementation for such un-initialization processes and routines in FIG. 4 will vary from client platform to client platform, as for the other routines discussed above. The structure, operation, etc. of such routines are well-known in the art, and they can be implemented using a number of fairly straightforward approaches without undue effort. Accordingly, they are not discussed in detail herein.

#### Description of Server Side System 180

##### Introduction

A high level flow diagram of the set of preferred processes implemented on server side system 180 of Natural Language Query System 100 is illustrated in FIGS. 11A through FIG. 11C. In a preferred embodiment, this process consists of a two step algorithm for completing the processing of the speech input signal, recognizing the meaning of the user's query, and retrieving an appropriate answer/response for such query.

The 1<sup>st</sup> step as illustrated in FIG. 11A can be considered a high-speed first-cut pruning mechanism, and includes the following operations: after completing processing of the speech input signal, the user's query is recognized at step 1101, so that the text of the query is simultaneously sent to Natural Language Engine 190 (FIG. 1) at step 1107, and to DB Engine 186 (also FIG. 1) at step 1102. By "recognized"

## US 7,050,977 B1

25

in this context it is meant that the user's query is converted into a text string of distinct native language words through the HMM technique discussed earlier.

At NLE 190, the text string undergoes morphological linguistic processing at step 1108: the string is tokenized the tags are tagged and the tagged tokens are grouped. Next the noun phrases (NP) of the string are stored at 1109, and also copied and transferred for use by DB Engine 186 during a DB Process at step 1110. As illustrated in FIG. 11A, the string corresponding to the user's query which was sent to the DB Engine 186 at 1102, is used together with the NP received from NLE 190 to construct an SQL Query at step 1103. Next, the SQL query is executed at step 1104, and a record set of potential questions corresponding to the user's query are received as a result of a full-text search at 1105, which are then sent back to NLE 190 in the form of an array at step 1106.

As can be seen from the above, this first step on the server side processing acts as an efficient and fast pruning mechanism so that the universe of potential "hits" corresponding to the user's actual query is narrowed down very quickly to a manageable set of likely candidates in a very short period of time.

Referring to FIG. 11B, in contrast to the first step above, the 2<sup>nd</sup> step can be considered as the more precise selection portion of the recognition process. It begins with linguistic processing of each of the stored questions in the array returned by the full-text search process as possible candidates representing the user's query. Processing of these stored questions continues in NLE 190 as follows: each question in the array of questions corresponding to the record set returned by the SQL full-text search undergoes morphological linguistic processing at step 1111: in this operation, a text string corresponding to the retrieved candidate question is tokenized, the tags are tagged and the tagged tokens are grouped. Next, noun phrases of the string are computed and stored at step 1112. This process continues iteratively at point 1113, and the sequence of steps at 1118, 1111, 1112, 1113 are repeated so that an NP for each retrieved candidate question is computed and stored. Once an NP is computed for each of the retrieved candidate questions of the array, a comparison is made between each such retrieved candidate question and the user's query based on the magnitude of the NP value at step 1114. This process is also iterative in that steps 1114, 1115, 1116, 1119 are repeated so that the comparison of the NP for each retrieved candidate question with that of the NP of the user's query is completed. When there are no more stored questions in the array to be processed at step 1117, the stored question that has the maximum NP relative to the user's query, is identified at 1117A as the stored question which best matches the user's query.

Notably, it can be seen that the second step of the recognition process is much more computationally intensive than the first step above, because several text strings are tokenized, and a comparison is made of several NPs. This would not be practical, nonetheless, if it were not for the fact that the first step has already quickly and efficiently reduced the candidates to be evaluated to a significant degree. Thus, this more computationally intensive aspect of the present invention is extremely valuable, however because it yields extremely high accuracy in the overall query recognition process. In this regard, therefore, this second step of the query recognition helps to ensure the overall accuracy of the system, while the first step helps to maintain a satisfactory speed that provides a real-time feel for the user.

26

As illustrated in FIG. 11C, the last part of the query/response process occurs by providing an appropriate matching answer/response to the user. Thus, an identity of a matching stored question is completed at step 1120. Next a file path corresponding to an answer of the identified matching question is extracted at step 1121. Processing continues so that the answer is extracted from the file path at 1122 and finally the answer is compressed and sent to client side system 150 at step 1123.

The discussion above is intended to convey a general overview of the primary components, operations, functions and characteristics of those portions of NLQS system 100 that reside on server side system 180. The discussion that follows describes in more detail the respective sub-systems.

#### 15 Software Modules used in Server Side System 180

The key software modules used on server-side system 180 of the NLQS system are illustrated in FIG. 5. These include generally the following components: a Communication module 500—identified as CommunicationServer ISAPI 500A (which is executed by SRE Server-side 182—FIG. 1 and is explained in more detail below), and a database process DBProcess module 501 (executed by DB Engine 186—FIG. 1). Natural language engine module 500C (executed by NLE 190—FIG. 1) and an interface 500B between the NLE process module 500C and the DBProcess module 500B. As shown here, CommunicationServerISAPI 500A includes a server-side speech recognition engine and appropriate communication interfaces required between client side system 150 and server side system 180. As further illustrated in FIG. 5, server-side logic of Natural Language Query System 100 also can be characterized as including two dynamic link library components: CommunicationServerISAPI 500 and DBProcess 501. The CommunicationServerISAPI 500 is comprised of 3 sub-modules: Server-side Speech Recognition Engine module 500A; Interface module 500B between Natural Language Engine modules 500C and DBProcess 501; and the Natural Language Engine modules 500C.

DB Process 501 is a module whose primary function is to connect to a SQL database and to execute an SQL query that is composed in response to the user's query. In addition, this module interfaces with logic that fetches the correct answer from a file path once this answer is passed to it from the Natural Language Engine module 500C.

#### 45 Speech Recognition Sub-System 182 on Server-Side System 180

The server side speech recognition engine module 500A is a set of distributed components that perform the necessary functions and operations of speech recognition engine 182 (FIG. 1) at server-side 180. These components can be implemented as software routines that are executed by server side 180 in conventional fashion. Referring to FIG. 4A, a more detailed break out of the operation of the speech recognition components 600 at the server-side can be seen as follows:

Within a portion 601 of the server side SRE module 500A, the binary MFCC vector byte stream corresponding to the speech signal's acoustic features extracted at client side system 150 and sent over the communication channel 160 is received. The MFCC acoustic vectors are decoded from the encoded HTTP byte stream as follows: Since the MFCC vectors contain embedded NULL characters, they cannot be transferred in this form to server side system 180 as such using HTTP protocol. Thus the MFCC vectors are first encoded at client-side 150 before transmission in such a way that all the speech data is converted into a stream of bytes



US 7,050,977 B1

27

without embedded NULL characters in the data. At the very end of the byte stream a single NULL character is introduced to indicate the termination of the stream of bytes to be transferred to the server over the INTERNET **160A** using HTTP protocol.

As explained earlier, to conserve latency time between the client and server, a smaller number of bytes (just the 13 MFCC coefficients) are sent from client side system **150** to server side system **180**. This is done automatically for each platform to ensure uniformity, or can be tailored by the particular application environment—i.e., such as where it is determined that it will take less time to compute the delta and acceleration coefficients at the server (26 mote calculations), than it would take to encode them at the client, transmit them, and then decode them from the HTTP stream. In general, since server side system **180** is usually better equipped to calculate the MFCC delta and acceleration parameters, this is a preferable choice. Furthermore, there is generally more control over server resources compared to the client's resources, which means that future upgrades, optimizations, etc., can be disseminated and shared by all to make overall system performance more reliable and predictable. So, the present invention can accommodate even the worst-case scenario where the client's machine may be quite thin and may just have enough resources to capture the speech input data and do minimal processing.

#### Dictionary Preparation & Grammar Files

Referring to FIG. 4A, within code block **605**, various options selected by the user (or gleaned from the user's status within a particular application) are received. For instance, in the case of a preferred remote learning system, Course, Chapter and/or Section data items are communicated. In the case of other applications (such as e-commerce) other data options are communicated, such as the Product Class, Product Category, Product Brand, etc. loaded for viewing within his/her browser. These selected options are based on the context experienced by the user during an interactive process, and thus help to limit and define the scope—i.e. grammars and dictionaries that will be dynamically loaded to speech recognition engine **182** (FIG. 1) for Viterbi decoding during processing of the user speech utterance. For speech recognition to be optimized both grammar and dictionary files are used in a preferred embodiment. A Grammar file supplies the universe of available user queries; i.e., all the possible words that are to be recognized. The Dictionary file provides phonemes (the information of how a word is pronounced—this depends on the specific native language files that are installed—for example, UK English or US English) of each word contained in the grammar file. It is apparent that if all the sentences for a given environment that can be recognized were contained in a single grammar file then recognition accuracy would be deteriorated and the loading time alone for such grammar and dictionary files would impair the speed of the speech recognition process.

To avoid these problems, specific grammars are dynamically loaded or actively configured as the current grammar according to the user's context, i.e., as in the case of a remote learning system, the Course, Chapter and/or Section selected. Thus the grammar and dictionary files are loaded dynamically according to the given Course, Chapter and/or Section as dictated by the user, or as determined automatically by an application program executed by the user.

The second code block **602** implements the initialization of Speech Recognition engine **182** (FIG. 1). The MFCC vectors received from client side system **150** along with the

28

grammar filename and the dictionary file names are introduced to this block to initialize the speech decoder.

As illustrated in FIG. 4A, the initialization process **602** uses the following sub-routines: A routine **602a** for loading an SRE library. This then allows the creation of an object identified as External Source with code **602b** using the received MFCC vectors. Code **602c** allocates memory to hold the recognition objects. Routine **602d** then also creates and initializes objects that are required for the recognition such as: Source, Coder, Recognizer and Results Loading of the Dictionary created by code **602e**, Hidden Markov Models (HMMs) generated with code **602f**, and Loading of the Grammar file generated by routine **602g**.

Speech Recognition **603** is the next routine invoked as illustrated in FIG. 4A, and is generally responsible for completing the processing of the user speech signals input on the client side **150**, which, as mentioned above, are preferably only partially processed (i.e., only MFCC vectors are computed during the first phase) when they are transmitted across link **160**. Using the functions created in External Source by subroutine **602b**, this code reads MFCC vectors, one at a time from an External Source **603a**, and processes them in block **603b** to realize the words in the speech pattern that are symbolized by the MFCC vectors captured at the client. During this second phase, an additional 13 delta coefficients and an additional 13 acceleration coefficients are computed as part of the recognition process to obtain a total of 39 observation vectors  $O_t$  referred to earlier. Then, using a set of previously defined Hidden Markov Models (HMMs), the words corresponding to the user's speech utterance are determined in the manner described earlier. This completes the word "recognition" aspect of the query processing, which results are used further below to complete the query processing operations.

It will be appreciated by those skilled in the art that the distributed nature and rapid performance of the word recognition process, by itself, is extremely useful and may be implemented in connection with other environments that do not implicate or require additional query processing operations. For example, some applications may simply use individual recognized words for filling in data items on a computer generated form, and the aforementioned systems and processes can provide a rapid, reliable mechanism for doing so.

Once the user's speech is recognized, the flow of SRE **182** passes to Un-initialize SRE routine **604** where the speech engine is un-initialized as illustrated. In this block all the objects created in the initialization block are deleted by routine **604a**, and memory allocated in the initialization block during the initialization phase are removed by routine **604b**.

Again, it should be emphasized that the above are merely illustrative of embodiments for implementing the particular routines used on a server side speech recognition system of the present invention. Other variations of the same that achieve the desired functionality and objectives of the present invention will be apparent from the present teachings.

#### Database Processor **186** Operation—DBProcess

Construction of an SQL Query used as part of the user query processing is illustrated in FIG. 4B, a SELECT SQL statement is preferably constructed using a conventional CONTAINS predicate. Module **950** constructs the SQL query based on this SELECT SQL statement, which query is used for retrieving the best suitable question stored in the database corresponding to the user's articulated query, (des-

## US 7,050,977 B1

29

ignated as Question here). A routine **951** then concatenates a table name with the constructed SELECT statement. Next, the number of words present in each Noun Phrase of Question asked by the user is calculated by routine **952**. Then memory is allocated by routine **953** as needed to accommodate all the words present in the NP. Next a word List (identifying all the distinct words present in the NP) is obtained by routine **954**. After this, this set of distinct words are concatenated by routine **955** to the SQL Query separated with a NEAR ( ) keyword. Next, the AND keyword is concatenated to the SQL Query by routine **956** after each NP. Finally memory resources are freed by code **957** so as to allocate memory to store the words received from NP for any next iteration. Thus, at the end of this process, a completed SQL Query corresponding to the user's articulated question is generated.

Connection to SQL Server—As illustrated in FIG. 4C, after the SQL Query is constructed by routine **710**, a routine **711** implements a connection to the query database **717** to continue processing of the user query. The connection sequence and the subsequent retrieved record set is implemented using routines **700** which include the following:

1. Server and database names are assigned by routine **711A** to a DBProcess member variable
2. A connection string is established by routine **711B**;
3. The SQL Server database is connected under control of code **711C**
4. The SQL Query is received by routine **712A**
5. The SQL Query is executed by code **712B**
6. Extract the total number of records retrieved by the query—**713**
7. Allocate the memory to store the total number of paired questions—**713**
8. Store the entire number of paired questions into an array—**713**

Once the Best Answer ID is received at **716** FIG. 4C, from the NLE **14** (FIG. 5), the code corresponding **716C** receives it passes it to code in **716B** where the path of the Answer file is determined using the record number. Then the file is opened **716C** using the path passed to it and the contents of the file corresponding to the answer is read. Then the answer is compressed by code in **716D** and prepared for transmission over the communication channel **160B** (FIG. 1).

#### NLQS Database **188**—Table Organization

FIG. 6 illustrates a preferred embodiment of a logical structure of tables used in a typical NLQS database **188** (FIG. 1). When NLQS database **188** is used as part of NLQS query system **100** implemented as a remote learning/training environment, this database will include an organizational multi-level hierarchy that consists typically of a Course **701**, which is made of several chapters **702**, **703**, **704**. Each of these chapters can have one or more Sections **705**, **706**, **707** as shown for Chapter 1. A similar structure can exist for Chapter 2, Chapter 3 . . . Chapter N. Each section has a set of one or more question—answer pairs **708** stored in tables described in more detail below. While this is an appropriate and preferable arrangement for a training/learning application, it is apparent that other implementations would be possible and perhaps more suitable for other applications such as e-commerce, e-support, INTERNET browsing, etc., depending on overall system parameters.

It can be seen that the NLQS database **188** organization is intricately linked to the switched grammar architecture described earlier. In other words, the context (or environment) experienced by the user can be determined at any moment in time based at the selection made at the section

30

level, so that only a limited subset of question-answer pairs **708** for example are appropriate for section **705**. This in turn means that only a particular appropriate grammar for such question-answer pairs may be switched in for handling user queries while the user is experiencing such context. In a similar fashion, an e-commerce application for an INTERNET based business may consist of a hierarchy that includes a first level "home" page **701** identifying user selectable options (product types, services, contact information, etc.), a second level may include one or more "product types" pages **702**, **703**, **704**, a third page may include particular product models **705**, **706**, **707**, etc., and with appropriate question-answer pairs **708** and grammars customized for handling queries for such product models. Again, the particular implementation will vary from application to application, depending on the needs and desires of such business, and a typical amount of routine optimization will be necessary for each such application.

#### Table Organization

In a preferred embodiment, an independent database is used for each Course. Each database in turn can include three types of tables as follows: a Master Table as illustrated in FIG. 7A, at least one Chapter Table as illustrated in FIG. 7B and at least one Section Table as illustrated in FIG. 7C.

As illustrated in FIG. 7A, a preferred embodiment of a Master Table has six columns—Field Name **701A**, Data Type **702A**, Size **703A**, Null **704A**, Primary Key **705A** and Indexed **706A**. These parameters are well-known in the art of database design and structure. The Master Table has only two fields—Chapter Name **707A** and Section Name **708A**. Both ChapterName and Section Name are commonly indexed.

A preferred embodiment of a Chapter Table is illustrated in FIG. 7B. As with the Master Table, the Chapter Table has six (6) columns—Field Name **720**, Data Type **721**, Size **722**, Null **723**, Primary Key **724** and Indexed **725**. There are nine (9) rows of data however, in this case,—Chapter\_ID **726**, Answer\_ID **727**, Section Name **728**, Answer\_Title **729**, PairedQuestion **730**, AnswerPath **731**, Creator **732**, Date of Creation **733** and Date of Modification **734**.

An explanation of the Chapter Table fields is provided in FIG. 7C. Each of the eight (8) Fields **720** has a description **735** and stores data corresponding to:

- AnswerID **727**—an integer that is automatically incremented for each answer given for user convenience
- Section\_Name **728**—the name of the section to which the particular record belongs. This field along with the AnswerID is used as the primary key
- Answer\_Title **729**—A short description of the title of the answer to the user query
- PairedQuestion **730**—Contains one or more combinations of questions for the related answers whose path is stored in the next column AnswerPath
- AnswerPath **731**—contains the path of a file, which contains the answer to the related questions stored in the previous column; in the case of a pure question/answer application, this file is a text file, but, as mentioned above, could be a multi-media file of any kind transportable over the data link **160**
- Creator **732**—Name of Content Creator
- Date\_of\_Creation **733**—Date on which content was created
- Date of Modification **734**—Date on which content was changed or modified

A preferred embodiment of a Section Table is illustrated in FIG. 7D. The Section Table has six (6) columns—Field

## US 7,050,977 B1

## 31

Name **740**, Data Type **741**, Size **742**, Null **743**, Primary Key **744** and Indexed **745**. There are seven (7) rows of data—Answer\_ID **746**, Answer\_Title **747**, PairedQuestion **748**, AnswerPath **749**, Creator **750**, Date of Creation **751** and Date of Modification **752**. These names correspond to the same fields, columns already described above for the Master Table and Chapter Table.

Again, this is a preferred approach for the specific type of learning/training application described herein. Since the number of potential applications for the present invention is quite large, and each application can be customized, it is expected that other applications (including other learning/training applications) will require and/or be better accommodated by another table, column, and field structure/hierarchy.

Search Service and Search Engine—A query text search service is performed by an SQL Search System **1000** shown in FIG. **10**. This system provides querying support to process full-text searches. This is where full-text indexes reside.

In general, SQL Search System determines which entries in a database index meet selection criteria specified by a particular text query that is constructed in accordance with an articulated user speech utterance. The Index Engine **1011B** is the entity that populates the Full-Text Index tables with indexes which correspond to the indexable units of text for the stored questions and corresponding answers. It scans through character strings, determines word boundaries, removes all noise words and then populates the full-text index with the remaining words. For each entry in the full text database that meets the selection criteria, a unique key column value and a ranking value are returned as well. Catalog set **1013** is a file-system directory that is accessible only by an Administrator and Search Service **1010**. Full-text indexes **1014** are organized into full-text catalogs, which are referenced by easy to handle names. Typically, full-text index data for an entire database is placed into a single full-text catalog.

The schema for the full-text database as described (FIG. **7**, FIG. **7A**, FIG. **7B**, FIG. **7C**, FIG. **7D**) is stored in the tables **1006** shown in FIG. **10**. Take for example, the tables required to describe the structure the stored question/answer pairs required for a particular course. For each table—Course Table, Chapter Table, Section Table, there are fields—column information that define each parameters that make up the logical structure of the table. This information is stored in User and System tables **1006**. The key values corresponding to those tables are stored as Full-Text catalogs **1013**. So when processing a full-text search, the search engine returns to the SQL Server the key values of the rows that match the search criteria. The relational engine then uses this information to respond to the query.

As illustrated in FIG. **10**, a Full-Text Query Process is implemented as follows:

1. A query **1001** that uses a SQL full-text construct generated by DB processor **186** is submitted to SQL Relational Engine **1002**.
2. Queries containing either a CONTAINS or FREETEXT predicate are rewritten by routine **1003** so that a responsive rowset returned later from Full-Text Provider **1007** will be automatically joined to the table that the predicate is acting upon. This rewrite is a mechanism used to ensure that these predicates are a seamless extension to a traditional SQL Server. After the compiled query is internally rewritten and checked for correctness in item **1003**, the query is passed to RUN TIME module **1004**. The function of module **1004** is to convert the rewritten

## 32

SQL construct to a validated run-time process before it is sent to the Full-Text Provider, **1007**.

3. After this, Full-Text Provider **1007** is invoked, passing the following information for the query:
  - a. A ft\_search\_condition parameter (this is a logical flag indicating a full text search condition)
  - b. A name of a full-text catalog where a full-text index of a table resides
  - c. A locale ID to be used for language (for example, word breaking)
  - d. Identities of a database, table, and column to be used in the query
  - e. If the query is comprised of more than one full-text construct; when this is the case Full-text provider **1007** is invoked separately for each construct.
4. SQL Relational Engine **1002** does not examine the contents of ft\_search\_condition. Instead, this information is passed along to Full-text provider **1007**, which verifies the validity of the query and then creates an appropriate internal representation of the full-text search condition.
5. The query request/command **1008** is then passed to Querying Support **1011A**.
6. Querying Support **1012** returns a rowset **1009** from Full-Text Catalog **1013** that contains unique key column values for any rows that match the full-text search criteria. A rank value also is returned for each row.
7. The rowset of key column values **1009** is passed to SQL Relational Engine **1002**. If processing of the query implicates either a CONTAINSTABLE( ) or FREETEXTTABLE( ) function, RANK values are returned; otherwise, any rank value is filtered out.
8. The rowset values **1009** are plugged into the initial query with values obtained from relational database **1006**, and a result set **1015** is then returned for further processing to yield a response to the user.

At this stage of the query recognition process, the speech utterance by the user has already been rapidly converted into a carefully crafted text query, and this text query has been initially processed so that an initial matching set of results can be further evaluated for a final determination of the appropriate matching question/answer pair. The underlying principle that makes this possible is the presence of a full-text unique key column for each table that is registered for full-text searches. Thus when processing a full-text search, SQL Search Service **1010** returns to SQL server **1002** the key values of the rows that match the database. In maintaining these full-text databases **1013** and full text indexes **1014**, the present invention has the unique characteristic that the full-text indices **1014** are not updated instantly when the full-text registered columns are updated. This operation is eliminated, again, to reduce recognition latency, increase response speed, etc. Thus, as compared to other database architectures, this updating of the full-text index tables, which would otherwise take a significant time, is instead done asynchronously at a more convenient time.

Interface Between NLE **190** and DB Processor **188**

The result set **1015** of candidate questions corresponding to the user query utterance are presented to NLE **190** for further processing as shown in FIG. **4D** to determine a “best” matching question/answer pair. An NLE/DBProcessor interface module coordinates the handling of user queries, analysis of noun-phrases (NPs) of retrieved questions sets from the SQL query based on the user query, comparing the retrieved question NPs with the user query NP, etc. between NLE **190** and DB Processor **188**. So, this part of the



US 7,050,977 B1

33

server side code contains functions, which interface processes resident in both NLE block **190** and DB Processor block **188**. The functions are illustrated in FIG. **4D**; As seen here, code routine **880** implements functions to extract the Noun Phrase (NP) list from the user's question. This part of the code interacts with NLE **190** and gets the list of Noun Phrases in a sentence articulated by the user. Similarly, Routine **813** retrieves an NP list from the list of corresponding candidate/paired questions **1015** and stores these questions into an (ranked by NP value) array. Thus, at this point, NP data has been generated for the user query, as well as for the candidate questions **1015**. As an example of determining the noun phrases of a sentence such as: "What issues have guided the President in considering the impact of foreign trade policy on American businesses?" NLE **190** would return the following as noun phrases: President, issues, impact of foreign trade policy, American businesses, impact, impact of foreign trade, foreign trade, foreign trade policy, trade, trade policy, policy, businesses. The methodology used by NLE **190** will thus be apparent to those skilled in the art from this set of noun phrases and noun sub-phrases generated in response to the example query.

Next, a function identified as Get Best Answer ID **815** is implemented. This part of the code gets a best answer ID corresponding to the user's query. To do this, routines **813A**, **813B** first find out the number of Noun phrases for each entry in the retrieved set **1015** that match with the Noun phrases in the user's query. Then routine **815a** selects a final result record from the candidate retrieved set **1015** that contains the maximum number of matching Noun phrases.

Conventionally, nouns are commonly thought of as "naming" words, and specifically as the names of "people, places, or things". Nouns such as John, London, and computer certainly fit this description, but the types of words classified by the present invention as nouns is much broader than this. Nouns can also denote abstract and intangible concepts such as birth, happiness, evolution, technology, management, imagination, revenge, politics, hope, cooker, sport, and literacy. Because of the enormous diversity of nouns compared to other parts of speech, the Applicant has found that it is much more relevant to consider the noun phrase as a key linguistic metric. So, the great variety of items classified as nouns by the present invention helps to discriminate and identify individual speech utterances much easier and faster than prior techniques disclosed in the art.

Following this same thought, the present invention also adopts and implements another linguistic entity—the word phrase—to facilitate speech query recognition. The basic structure of a word phrase—whether it be a noun phrase, verb phrase, adjective phrase—is three parts—[pre-Head string], [Head] and [post-Head string]. For example, in the minimal noun phrase—"the children," "children" is classified as the Head of the noun phrase. In summary, because of the diversity and frequency of noun phrases, the choice of noun phrase as the metric by which stored answer is linguistically chosen, has a solid justification in applying this technique to the English natural language as well as other natural languages. So, in sum, the total noun phrases in a speech utterance taken together operate extremely well as unique type of speech query fingerprint.

The ID corresponding to the best answer corresponding to the selected final result record question is then generated by routine **815** which then returns it to DB Process shown in FIG. **4C**. As seen there, a Best Answer ID **I** is received by routine **716A**, and used by a routine **716B** to retrieve an answer file path. Routine **716C** then opens and reads the answer file, and communicates the substance of the same to

34

routine **716D**. The latter then compresses the answer file data, and sends it over data link **160** to client side system **150** for processing as noted earlier (i.e., to be rendered into audible feedback, visual text/graphics, etc.). Again, in the context of a learning/instructional application, the answer file may consist solely of a single text phrase, but in other applications the substance and format will be tailored to a specific question in an appropriate fashion. For instance, an "answer" may consist of a list of multiple entries corresponding to a list of responsive category items (i.e., a list of books to a particular author) etc. Other variations will be apparent depending on the particular environment.

#### Natural Language Engine **190**

Again referring to FIG. **4D**, the general structure of NL engine **190** is depicted. This engine implements the word analysis or morphological analysis of words that make up the user's query, as well as phrase analysis of phrases extracted from the query.

As illustrated in FIG. **9**, the functions used in a morphological analysis include tokenizers **802A**, stemmers **804A** and morphological analyzers **806A**. The functions that comprise the phrase analysis include tokenizers, taggers and groupers, and their relationship is shown in FIG. **8**.

Tokenizer **802A** is a software module that functions to break up text of an input sentence **801A** into a list of tokens **803A**. In performing this function, tokenizer **802A** goes through input text **801A** and treats it as a series of tokens or useful meaningful units that are typically larger than individual characters, but smaller than phrases and sentences. These tokens **803A** can include words, separable parts of word and punctuation. Each token **803A** is given an offset and a length. The first phase of tokenization is segmentation, which extracts the individual tokens from the input text and keeps track of the offset where each token originated from in the input text. Next, categories are associated with each token, based on its shape. The process of tokenization is well-known in the art, so it can be performed by any convenient application suitable for the present invention.

Following tokenization, a stemmer process **804A** is executed, which can include two separate forms—inflectional and derivational, for analyzing the tokens to determine their respective stems **805A**. An inflectional stemmer recognizes affixes and returns the word which is the stem. A derivational stemmer on the other hand recognizes derivational affixes and returns the root word or words. While stemmer **804A** associates an input word with its stem, it does not have parts of speech information. Analyzer **806B** takes a word independent of context, and returns a set of possible parts of speech **806A**.

As illustrated in FIG. **8**, phrase analysis **800** is the next step that is performed after tokenization. A tokenizer **802** generates tokens from input text **801**. Tokens **803** are assigned to parts of a speech tag by a tagger routine **804**, and a grouper routine **806** recognizes groups of words as phrases of a certain syntactic type. These syntactic types include for example the noun phrases mentioned earlier, but could include other types if desired such as verb phrases and adjective phrases. Specifically, tagger **804** is a parts-of-speech disambiguator, which analyzes words in context. It has a built-in morphological analyzer (not shown) that allows it to identify all possible parts of speech for each token. The output of tagger **804** is a string with each token tagged with a parts-of-speech label **805**. The final step in the linguistic process **800** is the grouping of words to form phrases **807**. This function is performed by the grouper **806**,

US 7,050,977 B1

35

and is very dependent, of course, on the performance and output of tagger component **804**.

Accordingly, at the end of linguistic processing **800**, a list of noun phrases (NP) **807** is generated in accordance with the user's query utterance. This set of NPs generated by NLE **190** helps significantly to refine the search for the best answer, so that a single-best answer can be later provided for the user's question.

The particular components of NLE **190** are shown in FIG. **4D**, and include several components. Each of these components implement the several different functions required in NLE **190** as now explained.

Initialize Grouper Resources Object and the Library **900**—this routine initializes the structure variables required to create grouper resource object and library. Specifically, it initializes a particular natural language used by NLE **190** to create a Noun Phrase, for example the English natural language is initialized for a system that serves the English language market. In turn, it also creates the objects (routines) required for Tokenizer, Tagger and Grouper (discussed above) with routines **900A**, **900B**, **900C** and **900D** respectively, and initializes these objects with appropriate values. It also allocates memory to store all the recognized Noun Phrases for the retrieved question pairs.

Tokenizing of the words from the given text (from the query or the paired questions) is performed with routine **909B**—here all the words are tokenized with the help of a local dictionary used by NLE **190** resources. The resultant tokenized words are passed to a Tagger routine **909C**. At routine **909C**, tagging of all the tokens is done and the output is passed to a Grouper routine **909D**.

The Grouping of all tagged token to form NP list is implemented by routine **909D** so that the Grouper groups all the tagged token words and outputs the Noun Phrases.

Un-initializing of the grouper resources object and freeing of the resources, is performed by routines **909EA**, **909EB** and **909EC**. These include Token Resources, Tagger Resources and Grouper Resources respectively. After initialization, the resources are freed. The memory that was used to store all Noun Phrases are also de-allocated.

#### Additional Embodiments

In a e-commerce embodiment of the present invention as illustrated in FIG. **13**, a web page **1300** contains typical visible links such as Books **1310**, Music **1320** so that on clicking the appropriate link the customer is taken to those pages. The web page may be implemented using HTML, a Java applet, or similar coding techniques which interact with the user's browser. For example, if customer wants to buy an album C by Artist Albert, he traverses several web pages as follows: he first clicks on Music (FIG. **13**, **1360**), which brings up page **1400** where he/she then clicks on Records (FIG. **14**, **1450**). Alternatively, he/she could select CDs **1460**, Videos **1470**, or other categories of books **1410**, music **1420** or help **1430**. As illustrated in FIG. **15**, this brings up another web page **1500** with links for Records **1550**, with sub-categories—Artist **1560**, Song **1570**, Title **1580**, Genre **1590**. The customer must then click on Artist **1560** to select the artist of choice. This displays another web page **1600** as illustrated in FIG. **16**. On this page the various artists **1650** are listed as illustrated—Albert **1650**, Brooks **1660**, Charlie **1670**, Whyte **1690** are listed under the category Artists **1650**. The customer must now click on Albert **1660** to view the albums available for Albert. When this is done, another web page is displayed as shown in FIG. **17**. Again this web page **1700** displays a similar look and feel, but with the albums available **1760**, **1770**, **1780** listed under the heading Tides

36

**1750**. The customer can also read additional information **1790** for each album. This album information is similar to the liner notes of a shrink-wrapped album purchased at a retail store. One Album A is identified, the customer must click on the Album A **1760**. This typically brings up another text box with the information about its availability, price, shipping and handling charges etc.

When web page **1300** is provided with functionality of a NLQS of the type described above, the web page interacts with the client side and server side speech recognition modules described above. In this case, the user initiates an inquiry by simply clicking on a button designated Contact Me for Help **1480** (this can be a link button on the screen, or a key on the keyboard for example) and is then told by character **1440** about how to elicit the information required. If the user wants Album A by artist Albert, the user could articulate "Is Album A by Brooks available?" in much the same way they would ask the question of a human clerk at a brick and mortar facility. Because of the rapid recognition performance of the present invention, the user's query would be answered in real-time by character **1440** speaking out the answer in the user's native language. If desired, a readable word balloon **1490** could also be displayed to see the character's answer and so that save/print options can also be implemented. Similar appropriate question/answer pairs for each page of the website can be constructed in accordance with the present teachings, so that the customer is provided with an environment that emulates a normal conversational human-like question and answer dialog for all aspects of the web site. Character **1440** can be adjusted and tailored according to the particular commercial application, or by the user's own preferences, etc. to have a particular voice style (man, woman, young, old, etc.) to enhance the customer's experience.

In a similar fashion, an articulated user query might be received as part of a conventional search engine query, to locate information of interest on the INTERNET in a similar manner as done with conventional text queries. If a reasonably close question/answer pair is not available at the server side (for instance, if it does not reach a certain confidence level as an appropriate match to the user's question) the user could be presented with the option of increasing the scope so that the query would then be presented simultaneously to one or more different NLEs across a number of servers, to improve the likelihood of finding an appropriate matching question/answer pair. Furthermore, if desired, more than one "match" could be found, in the same fashion that conventional search engines can return a number of potential "hits" corresponding to the user's query. For some such queries, of course, it is likely that real-time performance will not be possible (because of the disseminated and distributed processing) but the advantage presented by extensive supplemental question/answer database systems may be desirable for some users.

It is apparent as well that the NLQS of the present invention is very natural and saves much time for the user and the e-commerce operator as well. In an e-support embodiment, the customer can retrieve information quickly and efficiently, and without need for a live customer agent. For example, at a consumer computer system vendor related support site, a simple diagnostic page might be presented for the user, along with a visible support character to assist him/her. The user could then select items from a "symptoms" page (i.e., a "monitor" problem, a "keyboard" problem, a "printer" problem, etc.) simply by articulating such symptoms in response to prompting from the support character. Thereafter, the system will direct the user on a

## US 7,050,977 B1

37

real-time basis to more specific sub-menus, potential solutions, etc. for the particular recognized complaint. The use of a programmable character thus allows the web site to be scaled to accommodate a large number of hits or customers without any corresponding need to increase the number of human resources and its attendant training issues.

As an additional embodiment, the searching for information on a particular web site may be accelerated with the use of the NLQS of the present invention. Additionally, a significant benefit is that the information is provided in a user-friendly manner through the natural interface of speech. The majority of web sites presently employ lists of frequently asked questions which the user typically wades item by item in order to obtain an answer to a question or issue. For example, as displayed in FIG. 13, the customer clicks on Help 1330 to initiate the interface with a set of lists. Other options include computer related items at 1370 and frequently asked questions (FAQ) at 1380.

As illustrated in FIG. 18, a web site plan for typical web page is displayed. This illustrates the number of pages that have to be traversed in order to reach the list of Frequently-Asked Questions. Once at this page, the user has to scroll and manually identify the question that matches his/her query. This process is typically a laborious task and may or may not yield the information that answers the user's query. The present art for displaying this information is illustrated in FIG. 18. This figure identifies how the information on a typical web site is organized: the Help link (FIG. 13, 1330) typically shown on the home page of the web page is illustrated shown on FIG. 18 as 1800. Again referring to FIG. 18, each sub-category of information is listed on a separate page. For example, 1810 lists sub-topics such as 'First Time Visitors', 'Search Tips', 'Ordering', 'Shipping', 'Tour Account' etc. Other pages deal with 'Account information' 1860, 'Rates and Policies' 1850 etc. Down another level, there are pages that deal exclusively with a sub—sub topics on a specific page such as 'First Time Visitors' 1960, 'Frequently Asked Questions' 1950, 'Safe Shopping Guarantee' 1940, etc. So if a customer has a query that is best answered by going to the Frequently Asked Questions link, he or she has to traverse three levels of busy and cluttered screen pages to get to the Frequently Asked Questions page 1950. Typically, there are many lists of questions 1980 that have to be manually scrolled through. While scrolling visually, the customer then has to visually and mentally match his or her question with each listed question. If a possible match is sighted, then that question is clicked and the answer then appears in text form which then is read.

In contrast, the process of obtaining an answer to a question using a web page enabled with the present NLQS can be achieved much less laboriously and efficiently. The user would articulate the word "Help" (FIG. 13, 1330). This would immediately cause a character (FIG. 13, 1340) to appear with the friendly response "May I be of assistance. Please state your question?". Once the customer states the question, the character would then perform an animation or reply "Thank you, I will be back with the answer soon". After a short period time (preferably not exceeding 5–7 seconds) the character would then speak out the answer to the user's question. As illustrated in FIG. 18 the answer would be the answer 1990 returned to the user in the form of speech is the answer that is paired with the question 1950. For example, the answer 1990: "We accept Visa, MasterCard and Discover credit cards", would be the response to the query 2000 "What forms of payments do you accept?"

Another embodiment of the invention is illustrated in FIG. 12. This web page illustrates a typical website that

38

employs NLQS in a web-based learning environment. As illustrated in FIG. 12, the web page in browser 1200, is divided into two or more frames. A character 1210 in the likeness of an instructor is available on the screen and appears when the student initiates the query mode either by speaking the word "Help" into a microphone (FIG. 2, 215) or by clicking on the link 'Click to Speak' (FIG. 12, 1280). Character 1210 would then prompt the student to select a course 1220 from the drop down list 1230. If the user selects the course 'CPlusPlus', the character would then confirm verbally that the course "CPlusPlus" was selected. The character would then direct the student to make the next selection from the drop-down list 1250 that contains the selections for the chapters 1240 from which questions are available. Again, after the student makes the selection, the character 1210 confirms the selection by speaking. Next character 1210 prompts the student to select 'Section' 1260 of the chapter from which questions are available from the drop down list 1270. Again, after the student makes the selection, character 1210 confirms the selection by articulating the 'Section' 1260 chosen. As a prompt to the student, a list of possible questions appear in the list box 1291. In addition, tips 1290 for using the system are displayed. Once the selections are all made, the student is prompted by the character to ask the question as follows: "Please ask your query now". The student then speaks his query and after a short period of time, the character responds with the answer preceded by the question as follows: "The answer to your question . . . is as follows: . . .". This procedure allows the student to quickly retrieve answers to questions about any section of the course and replaces the tedium of consulting books, and references or indices. In short, it is can serve a number of uses from being a virtual teacher answering questions on-the-fly or a flash card substitute.

From preliminary data available to the inventors, it is estimate that the system can easily accommodate 100–250 question/answer pairs while still achieving a real-time feel and appearance to the user (i.e., less than 10 seconds of latency, not counting transmission) using the above described structures and methods. It is expected, of course, that these figures will improve as additional processing speed becomes available, and routine optimizations are employed to the various components noted for each particular environment.

Again, the above are merely illustrative of the many possible applications of the present invention, and it is expected that many more web-based enterprises, as well as other consumer applications (such as intelligent, interactive toys) can utilize the present teachings. Although the present invention has been described in terms of a preferred embodiment, it will be apparent to those skilled in the art that many alterations and modifications may be made to such embodiments without departing from the teachings of the present invention. It will also be apparent to those skilled in the art that many aspects of the present discussion have been simplified to give appropriate weight and focus to the more germane aspects of the present invention. The microcode and software routines executed to effectuate the inventive methods may be embodied in various forms, including in a permanent magnetic media, a non-volatile ROM, a CD-ROM, or any other suitable machine-readable format. Accordingly, it is intended that the all such alterations and modifications be included within the scope and spirit of the invention as defined by the following claims.

What is claimed is:

1. A speech-enabled internet website operating on a server computing system and comprising:



US 7,050,977 B1

39

a receiving routine executing on the server computing system for receiving speech data associated with a user speech-based query, said speech data being characterized by a data content that is substantially inadequate by itself for permitting recognition of words articulated in said speech query; and

a speech recognition routine executing on the server computing system for completing recognition of said speech query using said speech data and said data content to generate a recognized speech query; and

a web page having a list of items, at least some of said list of items being selectable by a user based on said recognized speech query;

wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively.

2. The website of claim 1, wherein said web page displays an additional list of one or more items based on said recognized speech query.

3. The website of claim 1, wherein said website is adapted so that the user can navigate and locate information of interest using said speech query.

4. The website of claim 1, wherein said list of items include products and/or services offered by said website.

5. The website of claim 1, wherein said web page is implemented in HTML or as a Java applet.

6. The website of claim 1, wherein said website is further adapted to respond to a speech query concerning said list of items by returning a text or speech articulated response.

7. The website of claim 1, wherein said website is further adapted to interact on a real-time basis in response to one or more continuous speech queries.

8. The website of claim 1, wherein said speech recognition routine can complete recognition of said speech query with less latency than would that resulting if said additional data content were generated by a client platform used by the user.

9. The website of claim 1, wherein said data content constitutes a minimum amount of information that can be used by said speech recognition engine to complete accurate recognition of words and sentences in said speech query.

10. The website of claim 1, wherein the website also controls an interactive character agent presented to the user for assisting in handling said speech query.

11. The system of claim 10, wherein said interactive character agent provides suggestions for queries which the user can articulate.

12. The system of claim 10, wherein a different interactive character agent can be presented to different users providing speech utterances received by the server computing system.

13. The system of claim 10, wherein said interactive character agent is configured to perform a dialog of successive questions and answers with the user during an interactive session.

14. The system of claim 10, wherein said server computing system causes said interactive character agent to respond in real-time whenever the user provides selected speech input.

15. The website of claim 1, wherein said list of items correspond to topics associated with an interactive lesson tutorial.

40

16. The system of claim 1 wherein respective signal processing functions to be performed by the client platform and the server computing system are specified by an initialization routine.

17. The system of claim 16 wherein respective signal processing functions to be performed by the client platform and the server computing system are further specified in accordance with transmission characteristics associated with a communications channel used for said speech data.

18. The system of claim 1 wherein the server computing system is adapted to handle a client platform that can include a plurality of different hand held computing devices covering a range of differing respective computing capabilities.

19. The system of claim 1, wherein said speech data is formatted by a client device with at least one predetermined character used to designate end of an utterance.

20. The system of claim 19, wherein said predetermined character is a NULL character.

21. The system of claim 1, wherein the server computing system transfers speech related data for the web page using a hypertext transfer protocol (HTTP).

22. The system of claim 1, wherein a signal processing function performed by the client platform includes generating at least partial speech observation vectors using mel frequency cepstral coefficients.

23. The system of claim 1, wherein a signal processing function performed by the client platform includes at least calibrating speech and silence components of a speech utterance.

24. The system of claim 1 wherein the server computing system is further configured to perform a natural language processing operation on said recognized speech query to recognize a meaning of a sentence of words contained therein.

25. The system of claim 24 wherein said server computing system includes a plurality of separate natural language engines.

26. The system of claim 24 wherein said natural language processing operation is configured to compare a limited set of phrases from said recognized speech query with a separate set of phrases corresponding to predefined valid queries from users.

27. The system of claim 24 wherein text from said recognized speech query is presented to both a natural language engine for performing said natural language processing operation as well as to a database for identifying a meaning of said recognized speech query, such that a response can be provided by said database for at least some recognized speech queries before said natural language processing operation is completed.

28. The system of claim 24, further including a database query engine which performs part of said natural language operation by combining said speech query with search predicates to retrieve from a database a set of one or more potential responsive answers to said speech query.

29. The system of claim 1 wherein the server computing system is further configured to dynamically change a speech recognition grammar based on input provided by a user to selections available within said web page.

30. The system of claim 29 wherein multiple speech grammars are available and selectable within the web page, and such that speech input provided by the user for an item within the web page using a first grammar dynamically controls which one of a plurality of second grammars is loaded for speech recognition of subsequent speech input by the user.

## US 7,050,977 B1

41

31. The system of claim 29 wherein multiple speech grammars are selectable in a hierarchy within the web page, such that speech input provided by the user for an item within a first level menu of the web page using a first grammar dynamically controls which one of a plurality of second grammars at a second level menu of the web page and/or a second web page is loaded for speech recognition.

32. The system of claim 1 wherein the server computing system is further configured to dynamically change a speech recognition grammar based on spoken responses provided by a user during a real-time dialogue session conducted with an interactive electronic agent associated with said web page.

33. The system of claim 1, wherein said web page includes first tags which are selectable by one of a pointing device or a keyboard, and separate second tags selectable by speech input.

34. The system of claim 1, wherein said web page includes tags which can be selected by a pointing device and/or a keyboard and/or speech input.

35. The system of claim 1, wherein said web page and associated speech data is communicated to a client device using a hypertext transfer protocol (HTTP).

36. The system of claim 1, wherein said server computing system includes text to speech capability for outputting a response associated with said web page in audible form.

37. The system of claim 1, wherein said speech query is recognized by forming a concatenation of words and/or phrases derived from said speech query and using said concatenation as a search query for a database.

38. The system of claim 1, wherein the user can speak a help command while interacting with any web page maintained by the server computing system to cause an interactive character agent to appear.

39. A speech-enabled internet website operating on a server computing system and comprising:

- a receiving routine executing on the server computing system for receiving speech data associated with a user speech-based query, said speech data being characterized by a first data content that is substantially inadequate by itself for permitting recognition of words articulated in said speech query; and

- a speech recognition routine executing on the server computing system for completing recognition of said speech query using said speech data and said first data content to generate a recognized speech query; and

- a web page having a search engine for locating user selected information of interest, said search engine using a text query that is derived from said recognized speech query;

- wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively.

40. The website of claim 39, wherein said speech query is processed by more than one server computing system, so that multiple search engines are used for locating said information of interest.

41. The website of claim 39, wherein said web page includes a list of one or more items associated with assisting a user to diagnose a product or service problem, and which one or more items are also selectable by a user speech-based query.

42

42. The website of claim 39, wherein said website provides and controls an agent for assisting a user to interact with said website.

43. The website of claim 39, wherein said list of items correspond to topics associated with an interactive lesson tutorial.

44. A system for enabling a user web browser program to interact with a website using speech utterances, the system comprising:

- a receiving routine for receiving speech data associated with a speech utterance generated at a client platform, said speech data being characterized by a limited speech data content to reduce processing and transmission latencies; and

- a speech recognition routine executing on a server computing system for completing recognition of said speech utterance using said limited speech data content to generate a recognized speech query in real-time; and a web page routine for presenting one or more web pages to the user web browser program, wherein data content for said one or more web pages perceived by the user is controlled by said recognized speech query;

- wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively.

45. The system of claim 44, wherein said recognized speech query can include one of a number of predefined sentences recognizable by said system, and said speech query is recognized by identifying a candidate set of potential sentences from a number of predefined sentences, and then comparing each entry in the candidate set of potential sentences to said speech query to determine a matching recognized sentence.

46. The system of claim 45, wherein said speech utterance is compared against said candidate set of potential sentences by examining noun phrases.

47. The system of claim 45, wherein said candidate set of potential sentences are determined in part by a context dictionary loaded by said sentence recognition circuit in response to an operating environment presented by said system to a user.

48. The system of claim 44, wherein said speech utterance is processed by a natural language engine.

49. The system of claim 44, wherein environment variables experienced by the user within the web browser program are used for recognizing said speech query, such that said environmental variables vary in accordance with a web page being viewed by the user or a selection within a web page made by the user.

50. The website of claim 44, wherein said list of items correspond to topics associated with an interactive lesson tutorial.

51. The system of claim 44, wherein said limited speech data content does not include complete speech observation vectors which must be derived from said limited speech data content and input to said speech recognition routine before said speech utterance can be recognized.

52. The system of claim 44 wherein said limited speech data content comprises speech data that is transmitted continuously while the user is speaking and until silence is detected.

53. A method of interacting with a web-connected server using a client browser program, the method comprising the steps of:



## US 7,050,977 B1

43

- (a) receiving speech data associated with a speech utterance articulated by a user of the client platform, said speech data being characterized by a limited speech data content to reduce processing and transmission latencies; and
  - (b) completing recognition of said speech utterance using said limited speech data content to generate a recognized speech query at the web-connected server in real-time; and
  - (c) presenting one or more web pages to the user client web browser program, such that data content for said one or more web pages transmitted to the client browser program is controlled by said recognized speech query;
  - (d) allocating signal processing functions required to generate said recognized speech query between a client platform and the server computing system as needed based on computing resources available to said client and server computing systems respectively.
54. The method of claim 53 further including a step: performing a natural language processing operation to compare a limited set of phrases extracted from said recognized speech query with a separate set of phrases extracted from predefined valid queries from users.
55. The method of claim 53 further including a step: providing an interactive electronic character who provides suggestions for queries which the user can articulate.
56. The system of claim 55, further including a step: configuring said interactive character agent to engage in a dialog of successive questions and answers with the user during an interactive session.
57. The method of claim 53, further including a step: presenting an interactive character agent to the user in real-time in response to a spoken help command presented while interacting with any web page maintained by the server computing system.
58. The method of claim 53, further including a step: configuring said web page as a single page to a browser to allow a user to ask questions concerning any item identified in said database within said single page.
59. The method of claim 53, further including a step: forming a concatenation of words and/or phrases derived from said speech query and using said concatenation as a search query for a database.
60. A method of presenting information from a set of one or more web pages associated with a server interacting through a browser program with a client platform, the method comprising the steps of:
- (a) partially processing a speech utterance at the client platform to generate limited data content speech data, said limited data content speech data being configured to reduce processing and transmission latencies; and
  - (b) completing processing of said speech utterance using said limited speech data content to generate a recognized speech query at the server; and
  - (c) presenting content for the set of one or more web pages to the browser program, under control of said recognized speech query;
  - (d) allocating signal processing functions required to generate said recognized speech query between a client platform and the server computing system as needed based on computing resources available to said client and server computing systems respectively.
61. The method of claim 60, wherein said limited speech data content does not include complete speech observation vectors which must be derived from said limited speech data

44

content and input to said speech recognition routine before said speech utterance can be recognized.

62. The method of claim 60, wherein the server computing system transfers speech related data for the web page using a hypertext transfer protocol (HTTP).

63. The method of claim 60 further including a step: dynamically changing a speech recognition grammar based on input provided by a user to selections available within said web page.

64. A speech-enabled internet server computing system comprising:

- a receiving routine executing on the server computing system for receiving speech data associated with a user speech-based query, said speech data being characterized by a data content that is substantially inadequate by itself for permitting recognition of words articulated in said speech query; and

- a speech recognition routine executing on the server computing system for completing recognition of said speech query using said speech data and said data content to generate a recognized speech query;

wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively;

- a natural language routine executing on the server computing system and configured to process said recognized speech query to generate a natural language result in real-time;

- a web page having a list of items, at least some of said list of items being selectable by a user based on said natural language result;

- a database coupled to the server computing system for storing predefined answers which correspond to content for said list of items on said web page.

65. The speech-enabled internet server computing system of claim 64, wherein said web page contains links to other web pages which can be selected by speech queries.

66. The speech-enabled internet server computing system of claim 64, wherein said web page is a single page configured to allow a user to ask questions concerning any item identified in said database within said single page.

67. The speech-enabled internet server computing system of claim 64 wherein any and all of said list of items are selectable in a single screen.

68. The speech enabled internet server computing system of claim 67, wherein any and all of said list of items are selectable without scrolling through said web page.

69. A speech-enabled internet server computing system comprising:

- a receiving routine executing on the server computing system for receiving speech data associated with a user speech-based query, said speech data being characterized by a data content that is substantially inadequate by itself for permitting recognition of words articulated in said speech query; and

- a speech recognition routine executing on the server computing system for completing recognition of said speech query using said speech data and said data content to generate a recognized speech query;

wherein signal processing functions required to generate said recognized speech query can be allocated between a client platform and the server computing system as

## US 7,050,977 B1

45

needed based on computing resources available to said client platform and server computing system respectively;

a natural language routine executing on the server computing system and configured to process said recognized speech query to generate a natural language result based on an analysis of a selected limited set of phrases presented in said recognized speech query;

wherein said selected limited set of phrases are configured so that said natural language engine can generate said natural language result in real-time;

a web page having a list of items, at least some of said list of items being selectable by a user based on said natural language result;

a database coupled to the server computing system for storing content pertaining to said list of items on said web page.

70. The speech-enabled internet server computing system of claim 69 wherein said selected limited set of phrases include combinations of words and phrases contained in a grammar used by said speech recognition routine.

71. The speech-enabled internet server computing system of claim 69 wherein said selected limited set of phrases are generated dynamically from the recognized speech query.

72. The speech-enabled internet server computing system of claim 69 wherein said natural language engine compares said selected limited set of phrases to a set of phrases contained in predefined answers.

73. The speech-enabled internet server computing system of claim 69 wherein said natural language engine result is a single best answer.

74. A speech-enabled internet server computing system comprising:

a receiving routine executing on the server computing system for receiving speech data associated with a user speech-based query, said speech data being characterized by a data content that is substantially inadequate by itself for permitting recognition of words articulated in said speech query; and

a speech recognition routine executing on the server computing system for completing recognition of said speech query using said speech data and said data content to generate a recognized speech query;

wherein signal processing functions required to generate said recognized speech query can be allocated

46

between a client platform and the server computing system as needed based on computing resources available to said client platform and server computing system respectively;

a natural language routine executing on the server computing system and configured to process said recognized speech query to generate a natural language result based on an analysis of a selected limited set of phrases presented in said recognized speech query;

wherein said selected limited set of phrases are configured so that said natural language engine can generate said natural language result and a response can be provided to said user speech-based query in real-time;

a web page having a list of items, at least some of said list of items being selectable by a user based on said natural language result;

a database coupled to the server computing system for storing content pertaining to said list of items on said web page;

an electronic conversational agent adapted to interact with the user and mimic behavior of a human agent through a native language interactive real-time dialog session with the user.

75. The speech-enabled internet server computing system of claim 74, wherein said electronic conversational agent is presented within a client browser.

76. The speech-enabled internet server computing system of claim 74, wherein said electronic conversational agent is a visual character on a screen.

77. The speech-enabled internet server computing system of claim 74, wherein said electronic conversational agent is configured to articulate suggestions to the user for appropriate speech queries.

78. The speech-enabled internet server computing system of claim 74, wherein said electronic conversational agent is adapted to have configurable perception parameters which are adjusted and tailored to said content pertaining to said list of items.

\* \* \* \* \*

1 R. Joseph Trojan CA Bar No. 137,067  
trojan@trojanlawoffices.com  
2 TROJAN LAW OFFICES  
9250 Wilshire Blvd., Suite 325  
3 Beverly Hills, CA 90212  
Telephone: (310) 777-8399  
4 Facsimile: (310) 777-8348

5 Attorneys for Plaintiff,  
PHOENIX SOLUTIONS, INC.

6  
7  
8 **UNITED STATES DISTRICT COURT**  
9 **NORTHERN DISTRICT OF CALIFORNIA**  
10 **SAN FRANCISCO**

11  
12 PHOENIX SOLUTIONS, INC., a  
California corporation,

13  
14 Plaintiff,

15  
16 v.

17  
18 WELLS FARGO BANK, N.A., a  
Delaware corporation,

19  
20 Defendant.

CASE NO. CV08-0863 MHP

**PLAINTIFF PHOENIX SOLUTIONS,  
INC.'S [PROPOSED] ORDER  
STRIKING DEFENDANT'S  
AFFIRMATIVE DEFENSE NUMBERS  
35, 36, 37 AND 38 UNDER FEDERAL  
RULE OF CIVIL PROCEDURE, RULE  
12(f)**

DATE: July 14, 2008

TIME: 2:00 p.m.

JUDGE: MARILYN H. PATEL

21  
22  
23  
24  
25  
26 //

27 //

28 **[PROPOSED] ORDER**

**CV08-0863 MHP**

1 This matter came before the Court on motion brought by Plaintiff to strike  
2 Defendant's affirmative defense numbers 35, 36, 37 and 38. A hearing for  
3 Plaintiff's Motion to Strike was heard before this Court on Monday, July 14, 2008,  
4 at 2:00 p.m. Plaintiff Phoenix Solutions, Inc., ("Phoenix") and Defendant Wells  
5 Fargo Bank, N.A., ("Wells Fargo") were both represented by counsel in this matter  
6 and oral arguments were heard on this same day. This Court, having considered all  
7 papers and oral arguments in support thereof, and in opposition thereto, for good  
8 cause appearing, and all other matters of record presented before the Court or to  
9 which this Court may take judicial notice, does HEREBY GRANT THE MOTION  
10 AND ORDERS AS FOLLOWS:

11  
12 THIS COURT HEREBY STRIKES DEFENDANT'S AFFIRMATIVE  
13 DEFENSE NUMBERS 35, 36, 37 AND 38 UNDER FEDERAL RULES OF  
14 CIVIL PROCEDURE, RULE 12(f).

15  
16  
17  
18  
19  
20  
21  
22  
23  
24 Dated: \_\_\_\_\_

\_\_\_\_\_  
Hon. Judge Marilyn H. Patel  
United States District Court  
Northern District of California

25  
26  
27  
28 [PROPOSED] ORDER

CV08-0863 MHP